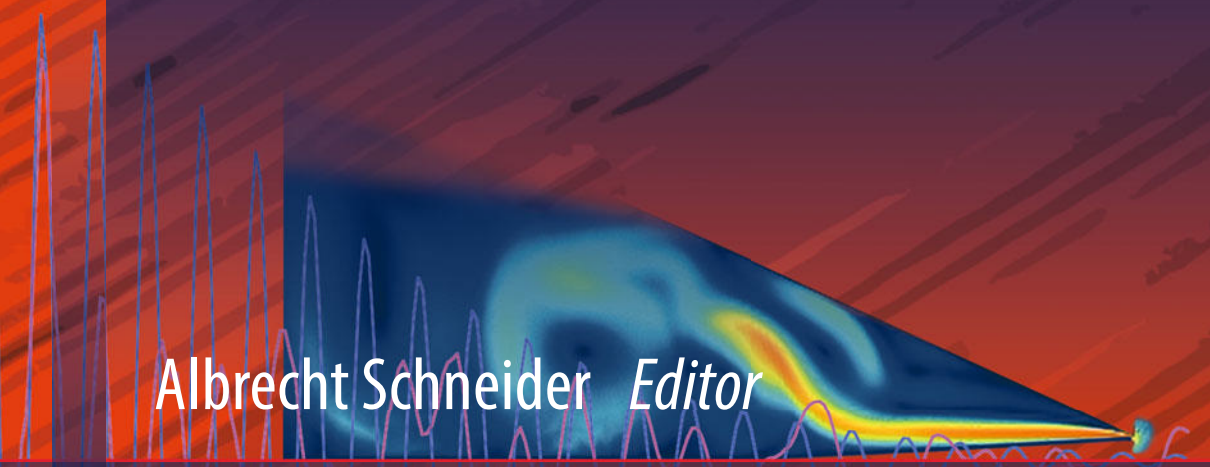
Current Research in Systematic Musicology



Albrecht Schneider *Editor*

Studies in Musical Acoustics and Psychoacoustics

Springer

Current Research in Systematic Musicology

Volume 4

More information about this series at http://www.springer.com/series/11684

Albrecht Schneider
Editor

Studies in Musical Acoustics and Psychoacoustics

Springer

Editor
Albrecht Schneider
Institut für Systematische
Musikwissenschaft
Universität Hamburg
Hamburg
Germany

ISSN 2196-6966 ISSN 2196-6974 (electronic)
Current Research in Systematic Musicology
ISBN 978-3-319-47291-1 ISBN 978-3-319-47292-8 (eBook)
DOI 10.1007/978-3-319-47292-8

Library of Congress Control Number: 2016953316

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Contents

Introduction

In the volume at hand, topics in musical acoustics and perception of sound are treated from a range of perspectives and with various methods. In general, the scientific field of musical acoustics is structured into several areas, some of which are close to physics, while others relate to music and musicology as well as to disciplines engaged in the study of sensation and perception. Musical instruments and the voice (of both humans and other species) are studied in regard to sound production and radiation of sound from a source into the environment. Sound production mechanisms often account also for pitch structures and timbral qualities available from individual instruments or from 'families' of instruments. Room acoustics is needed to understand the radiation processes including reflection and refraction of sound waves at boundaries as well as dissipation of sound energy within specific geometries.

Musical sound is produced, by musicians as well as singers, with the aim of communicating with a (real or virtual) listener. Of course, the player of an instrument or a singer acts himself or herself as a listener and makes use of his or her analytical listening capability to control, first of all, the parts (muscles, tendons, etc.) of his or her body involved, as effectors, in the production of sound. Playing an instrument or singing thus is based on feedback loops which control sound production, pitches and intonation as well as timbre and dynamic parameters in a musical performance.

Ideally, music as performed in a live event like a concert addresses an audience of appreciative subjects, meaning subjects capable of perceiving music as textures of sound from which the structure of a composition or improvisation may be gathered. Musical appreciation, however, can be viewed as the terminal point of a process that starts with sensation of sounds at the ears as the relevant peripheral sense organs (there are indications that also the vestibular system may be excited by very loud sounds). Taking the peripheral auditory system as a first stage of 'information pickup' and signal analysis, further analysis of sound in regard to salient features and pattern recognition is conducted along the auditory pathway and, finally, in cortical areas of the brain. Though there seems to be a hierarchy from initial sensation (which must be fast to allow for real-time processing

of complex sounds as well as efferent feedback activation within the auditory system) to perception directed to salient features and pattern recognition, followed by an evaluation of sensory input in cortical networks that might yield 'auditory objects', it is in fact the structure of the sound signal and the anatomical and physiological organization of the inner ear and the auditory pathway that determine perception of pitch, timbre and loudness. In this respect, a bottom-up approach to sound and music perception based on musical acoustics and psychoacoustics seems necessary notwithstanding the obvious role of musical training and sociocultural factors which can shape perception and cognition of music in individuals.

From what has been sketched in the preceding paragraph, one may view musical acoustics as centred on musical instruments in regard to mechanisms of sound production and radiation, but also including properties governing pitch, timbre and dynamic structures that in turn are relevant for sensation and perception of musical sound. Furthermore, studying actual playing and singing techniques can give insight into functional aspects of sound production and musical expression. However, musical acoustics includes also the formation of tone systems as well as scales, tunings and intonation patterns. Furthermore, while physical acoustics (traditionally a part of mechanics) may be conceived as a fundamental science treating the theory of vibration and sound with little regard to actual sensation and perception, musical acoustics relates to sensation and perception as well as to the production of sound in mammalian or other species in many ways. It is from such an integrative perspective that perceptual aspects and results from experiments involving musicians and/or listeners will be considered within the broader area of musical acoustics.

Several articles in this volume deal with the acoustics and organology of peculiar instruments as well as with certain types of instruments. Shigeru Yoshikawa offers a comprehensive study on Japanese flutes with a focus on their construction and acoustic properties as well as on playing techniques (such as cross-fingerings needed to produce a variety of pitches) as a factor that conditions intonation and timbral qualities. His article includes the classical and the modern shakuhachi, the nohkan (a transverse bamboo flute) and the shinobue (another transverse bamboo flute). Starting from the structural properties of each instrument (such as the shape of the embouchure and the bore), Yoshikawa has calculated admittance and resonance conditions in relation to fingerings. Also, he discusses the data obtained from a number of experiments including measurements as well as sound analyses. Taken together, the empirical evidence shows that the Japanese flute types studied in this article differ from European flutes in several respects, among them construction and materials, but most significantly sound properties which feature wind noise from blowing as an essential component of the sound. As Yoshikawa concludes, Japanese flutes are constructed for producing distinct timbral qualities with an emphasis on spectral energy in higher-frequency bands.

The study of the Chinese Qin carried out by Chris Waltham, Kimi Coaldrake, Evert Koster and Yang Lan provides fresh information from current research on the acoustics of an instrument that has a long history and is highly regarded in Chinese music tradition. The Qin is one of several plucked zither types of East Asia which

are of interest in regard to their construction, materials and sound properties. The acoustics of the Qin (of which little was known so far) is investigated, by Waltham and coworkers, by vibroacoustical measurement as well as a FEM modelling approach. The article offers empirical data in regard to materials, vibroacoustics, sound analysis and the FEM model chosen for this study.

Florian Pfeifle and Malte Münster have studied sound generation in two instruments widely used in rock and pop music genres, the Wurlitzer E-piano, and the Fender-Rhodes E-piano. While the Rhodes employs an electromechanical set-up for the generation and pickup of sound, the Wurlitzer uses electrostatical effects. Pfeifle and Münster have measured the vibrational patterns of sound generating elements (tine, bar, reed) with a high-speed camera and have made analyses of the electrical properties of the pickup systems as well as of the actual sound produced so that the mechanical and the electronic data form the basis, as intermediate results, for a finite-element modelling (FEM) and finite difference calculation approach to finding characteristics of sound generation in the two instruments. The article shows that the peculiar timbre in both instruments is largely due to the specific set-up and geometry of their respective pickup systems.

Jost Leonhardt Fischer investigates the feedback of different room geometries on the sound radiated from an organ pipe. Previous studies have demonstrated that pipes being placed on the same wind chest can influence each other because of acoustic coupling. In addition, one needs to consider sound radiation from individual pipes being hampered by the presence of several or even many pipes in their immediate surrounding as well as by structural parts of the organ (such as beams or brackets). Applying numerical simulation methodology, Fischer shows that sound waves radiated from an organ pipe undergo significant variation in regard to frequencies and amplitudes depending on the geometry of the reflecting surface. The effect is particularly visible if a pipe is located inside a closed swell chamber.

Shigeru Yoshikawa and Yu Nobara address acoustical problems associated with mutes as are used in playing brass instruments such as the French horn and the trumpet. In particular, they consider the stopping and straight mutes for the horn and the straight, cup and wah-wah mutes for the trumpet. From modelling the horn and the trumpet on the basis of branching theory and from extensive numerical calculation including transmission matrix (T-matrix) representation of the horn system as well as from data obtained in their own measurements, Yoshikawa and Nobara discuss acoustical parameters such as input impedance and admittance, internal pressure distribution in the bore and transmission function. Among their explanations of the effects mutes have for changes in resonance frequencies, modes and spectral energy distribution is that hand-stopping, in the French horn, causes a *descent* in pitch (while mutes in general sharpen pitches).

Malte Kob surveys a number of factors relevant for damping in musical instruments as well as parameters and methods suited to measuring damping in a vibrating system. Among the approaches that have been taken in experiments on musical instruments, one finds measurement of the loss factor, of the reverberation time (T_{60}), or of the −3dB bandwidth, respectively. In this article, results obtained from the measurement of vibrational patterns of a metal tongue are presented in a

comparative perspective. The study was undertaken using the reverberation time method in the time domain and the −3dB method in the frequency domain.

James Beauchamp bases his comparative study of vocal and violin vibrato viewed in regard to the source/filter model (prominent in phonetics but also in instrument acoustics) on signal processing methodology. In particular, he uses a range of tools available from the sndan package developed by Beauchamp and Maher. Applying the source/filter model to the analysis of complex sounds means one needs to separate the source waveform and spectrum from the filter defined by its transfer function. In complex sounds such as musical tones sung with vibrato in bel canto operatic style or played on a violin with vibrato (a nearly periodic change of the length of a vibrating string affected by finger movements), partial frequencies and amplitudes are modulated more or less sinusoidally, causing time-variant spectra. One method to track partials undergoing such modulation is the McAulay–Quatieri peak-picking algorithm (implemented as part of sndan). With detailed signal analyses, Beauchamp demonstrates how to separate the source spectrum for sung vibrato tones and for glides in violin tones (even though the source spectrum for such violin glides itself varies considerably with time).

Robert Mores proposes vowel quality (VQ) as a descriptor for the timbre of sounds recorded from Italian masterpieces (mostly Stradivari and Guarneri violins). VQ describes a vowel as produced by the human vocal tract by two key parameters, tongue backness and tongue height. The methodology outlined and discussed in this study is based on voice analysis and signal processing. Taking similarities in the sound structure of vowels and violin sounds as a starting point, formants are extracted and VQ parameters are calculated from recordings of violin tones in an automated process. Results are matched to the IPA chart of vowels and are validated by behavioural experiments in which subjects had to select voice sounds of different VQ so as to match violin sounds in terms of the VQ. Mores' study shows that VQ is an appropriate descriptor of violin timbre suited to be used in a comparative analysis of sounds recorded from a range of historic or contemporary violins.

Albrecht Schneider and Marc Leman investigate sound characteristics and the tuning of a historic carillon, founded by Joris Dumery for the city of Bruges (Flanders), in the eighteenth century. This carillon today comprises 47 bells, of which 26 are from Dumery. The bells were recorded on the belfry before a restoration took place lately. Though the daily use of the carillon might affect material and acoustical properties of its bells to some degree, the carillon could still be investigated for its original tuning. Since the spectra from bell sounds contain components which conform to segments of a harmonic series as well as numerous inharmonic components, pitch perception from bell sounds often is ambiguous. The article discusses the phenomenon of the so-called strike note in bell sounds and concepts of virtual pitch in regard to pitch perception of inharmonic sounds. Data from some recent behavioural experiments on pitch ambiguity employing the sounds from the Dumery carillon as stimuli are also included into this study.

Tim Ziemer's article on source width in music production (recording, mixing, mastering) comprises both theoretical modelling and experimental work. As is

known from acoustical measurements, the actual sound radiation patterns of individual instruments relate to their geometry as well as to the register and the dynamic level on which tones are played. Hence, the physical sound source can vary on several parameters. In addition, room acoustics can influence the perception listeners may have of a source in regard to spatial attributes. Different recording and mixing techniques either may preserve spatial characteristics of sound sources (e.g. natural voices or instruments), or may deliberately change the apparent source width for listeners (e.g. by applying stereophonic effects to monaural signals). After providing the fundamental concepts of room acoustics, Ziemer discusses source width in music production and recording with respect to stereo and surround set-ups as well as in ambisonics and in wave field synthesis. Finally, he reports an experiment in which sounds recorded from various instruments with a microphone array were projected to a large number of virtual listening positions with the aim of finding relations between sound field parameters and apparent source width.

Christiane Neuhaus reviews methods in neuromusicology (aka cognitive neuroscience of music) such as transcranial magnetic stimulation (TMS), functional magnetic resonance imaging (fMRI), positron-emission tomography (PET), electroencephalography (EEG) and measurement of event-related potentials (ERPs). Studies of neural processes and functions in the brain in regard to sensation and perception of sound and music were begun decades ago with EEG methodology but gained new impetus when imaging techniques such as PET and fMRI became available. Though research on 'the musical brain' employing the aforenamed technologies has largely expanded the scope of topics and methods known from 'classical' psychoacoustics and music psychology, there are also constraints that must be taken into account. In effect, each method needs to be evaluated, on theoretical grounds as well as with respect to results obtained from experiments or simulations.

Jonas Braasch, Selmer Bringsjord, Nikhil Deshpande, Pauline Oliveros and Doug Van Nort report on the current state of an ongoing project labelled *Creative Artificially-Intuitive and Reasoning Agent*, CAIRA, an intelligent music system capable to be used in performances of music from various genres, among them traditional and free jazz. The system has a dual architecture and combines signal-driven bottom-up analysis with tools from computational auditory scene analysis (CASA) and logic-based top-down reasoning. The bottom-up analysis is directed to psychoacoustic parameters and uses auditory models for pitch, timbre and loudness, while modules in addition extract parameters relating to sonic textures and gestures. The concept of an intelligent agent such as in CAIRA is that it needs to 'understand' the essential features of the genres and the style that is played by a musician (with whom the agent interacts in a live situation, that is, in real time). This implies the system needs to incorporate modules comprising algorithms for machine learning (as in this system).

The article on keyboard temperaments from Albrecht Schneider and the late Andreas Beurmann offers a historical review of sources from music theory and organology relating to tone systems, tunings and temperaments as well as empirical data from measurements and behavioural experiments. Since just intervals form the

basis of most tone systems and scale types (for acoustic and psychoacoustic reasons), the number of pitches (m) extends to $m > 12$ if chords in various major and minor keys shall be played in just intonation. Regarding technical limitations that existed for mechanical keyboard instruments as well as for musical performance practice, temperaments and actual tunings had to reduce the large number of tones and pitches resulting from modal and chordal structures conceived in just intervals to a much smaller number $n = 12$ of keys available in a conventional keyboard. Temperaments calculated and proposed by Werckmeister or Vallotti can be viewed as a compromise in regard to tuning scales and intervals so that with only 12 tones/pitches to the octave, modulation through a range of keys is possible, while beats and roughness are kept within certain limits. The article presents data for Werckmeister III and Vallotti tuned on a historical Kirckman harpsichord. Also included are data from behavioural experiments where subjects had to rate musical excerpts presented in various tunings.

Hamburg Albrecht Schneider
July 2016

Japanese Flutes and Their Musical Acoustic Peculiarities

Shigeru Yoshikawa

Abstract Representative Japanese bamboo flutes, the *shakuhachi*, *nohkan*, and *shinobue* are investigated from musical acoustic viewpoint. The end-blown longitudinal flute, shakuhachi has only five tone holes, and several cross fingerings causes pitch sharpening (called *intonation anomaly*) as well as characteristic timbre, particularly in the second and third registers. Also, acoustical differences between classical and modern shakuhachis are made clear. The nohkan has a special tube device, "throat" (called *nodo* in Japanese), which is inserted between the embouchure hole and the top tone hole to narrow the bore. This throat significantly upsets the expected octave relation between the first and second registers. The octave is enlarged for low-pitched fingerings, while it is strongly shrunk for high-pitched fingerings. The nohkan is compared with the piccolo concerning an interesting fingering with two extremely distant open tone holes. The upper tone hole functions as an octave hole. The shinobue has another special device, a membrane hole over which the inner skin of the bamboo node (called *chikushi* in Japanese) is glued. The membrane vibration driven by the bore resonance pressure produces brilliant and distinctive sounds due to the resulting high-frequency emphasis. These unique structural properties of Japanese flutes bring about their musical and acoustical peculiarities not usually observed in Western flutes.

1 Introduction

Traditional lip-driven brass instruments do not exist in Japan as well as in Asia in contrast with many brass instruments in the West. On the other hand, a variety of woodwind instruments made of bamboo have been played in Japan, Korea, and China. Particularly, there are flute-type instruments in wide varieties in Japan. Generally, they are called *fue* (as a suffix, *-bue*, e.g. *yokobue*, which is general term

S. Yoshikawa (✉)
1-27-22 Aoyama, Dazaifu 818-0121, Japan
e-mail: shig@lib.bbiq.jp

A. Schneider (ed.), *Studies in Musical Acoustics and Psychoacoustics*,
Current Research in Systematic Musicology 4,
DOI 10.1007/978-3-319-47292-8_1

for transverse flutes) including the end-blown longitudinal bamboo flute, *shakuhachi*.

The objective of this chapter is to explore the musical acoustics of Japanese flutes (*fue*) while considering distinctive characteristics in Asian music. It will be demonstrated that structural peculiarities of Japanese flutes bring about their musical peculiarities. We should pay our attention onto the embouchure edge, mouth-hole geometry, finger-hole geometry, etc., which have deep relation with their sounds.

The origin of *fue* might be considered as *iwabue* (stone whistle). Typical stone whistles excavated from several remains in the *Johmon* period (around BC 3000) are made of natural stone with the Y-shaped open holes (stone shape and size are various, 10 cm × 6 cm in rough average). If the branched Y-shaped holes are closed/opened by fingers when it is blown from the bottom hole, very vivid, clear, and powerful tones (with nearly sinusoidal waveforms of pitches around C_7, D_7, and E_7) are generated [1, 2]. These iwabue tones seem to have created tonal sensations toward *fue* for the Japanese from the ancient to the modern.

In this chapter the *shakuhachi*, *nohkan*, and *shinobue* are considered from the viewpoint of musical acoustics. The roots of these flue instruments were imported from China in its classic period of *Tang* dynasty (618–690, 705–907) and in Japanese periods of *Asuka* (538–710) and *Nara* (710–794). However, these three instruments as well as others are completely changed to Japanese instruments based on their tonal sensations mentioned above. The shakuhachi was fue for the *komuso* (wandering monks) in the *Edo* period (1603–1867), the nohkan was fue for the *samurai* (faithful warriors) in the *Muromachi* period (1334–1573), and the shinobue was fue for the common people in the Edo period. Their tones have taken on unique characteristics born from their histories and relations with the society.

2 The Shakuhachi

According to Malm [3], "One of the easiest ways to approach the music of another culture is through its flute literature. There seems to be something in the tone of the flute that has a universal appeal. This catholic quality is amply illustrated by the example of the *shakuhachi*."

2.1 Brief History

The shakuhachi was originally introduced to Japan from China in the Tang dynasty around 750. Since this ancient shakuhachi has been preserved in the *Shosoin* warehouse of the *Tohdaiji* temple, it is called the *Shosoin shakuhachi* (its

musicological term is the *gagaku shakuhachi*). This shakuhachi, which had six tone holes to play a Chinese diatonic scale (e.g. D-E-Gb-G-A-B-D), was adapted to play a Japanese pentatonic scale (D-E-G-A-B-D) by removing the second (counted from the bottom) tone hole (Gb) around early 16th century. Moreover, the positions of five tone holes were modified to make effective use of pitch bending (e.g. Eb and A#) by the *meri/kari* blowing (by pulling down/up player's jaw) and by half-covering the tone hole(s) around the 17th century, and thus a scale pattern D-F-G-A-C-D was established when playing the shakuhachi with the standard length of *ichi* (one) *shaku* and *hachi* (eight) *sun* (54.5 cm) [4–6].

Since this shakuhachi (made from the root end of bamboo) was played exclusively by a group of wandering priests (called *komuso* having faith in *Fukeshu*, a sect of Buddism), it is called the *komuso* (or *Fuke*) *shakuhachi* and regarded as the origin of the modern shakuhachi [3, 5, 6]. The history of changes from the Shosoin shakuhachi to the komuso shakuhachi is very complicated and indefinite [3–5]. The former was probably played in court chamber music (*gagaku*); the latter was played in solo.

In 1871 (the 4th year of *Meiji*) Fukeshu was abolished from various reasons under the Meiji Restoration which executed strong national policy of the westernization. The Western-oriented music was eagerly promoted; the Japanese traditional music was coldly shunned. In this early Meiji period the shakuhachi became open to common people and was used in ensemble music with string instruments such as the *soh* (*koto*) and the *shamisen* (three-stringed instruments). In the late 20th century non-Japanese performers and makers of the shakuhachi appeared in the West as well as in Japan. Nowadays the International Shakuhachi Festival has been held every a few years. The shakuhachi is an international musical instrument with its contemporary vitality.

2.2 *Unique Structural Properties*

As properly pointed out by Malm [3], the characteristic properties of the shakuhachi are (1) the oblique blowing edge, (2) only five tone holes (four on the front and one on the back), and (3) the inner bore geometry from the edge to the root bottom. In this section acoustical effects of these properties will be demonstrated. Before that, unique these properties are to be explained in more detail.

Embouchure edge: Using his Fig. 23, Malm [3] described the evolution of the edge shape of the end-blown instruments as follows: The original pipe was merely blown across the top end just as children do a hollow bottle. The Chinese end-blown instrument *dungxiao* (in Japanese, *dohsho*), which has been considered as the origin of the shakuhachi, has the edge obliquely cut inward. However, the shakuhachi is unique in the way in which its embouchure is constructed [3]. It is cut outward, the exact opposite of the Chinese manner. This should be a Japanese innovation.

Two examples of the shakuhachi edge are shown in Fig. 1. A shallow edge with a short cut is depicted in Fig. 1a; a deep edge with a long cut in Fig. 1b. The edge shape and geometry are very essential to the players because the embouchure edge is the point joining the instrument and player. The starting transient of a tone largely depends on the embouchure edge. The shape of back side, on which player's lower lip is placed, is also important when the meri/kari blowing is applied. It is different from each other as shown in Fig. 1a, b.

The edge shape and geometry decisively determine the harmonic generation of the shakuhachi sound [7–9]. The blowing edge forms the source spectrum through the interaction with the air flow from the player. The source spectrum is then modified by the resonance characteristics of the bore and by the radiation characteristics of end openings at the embouchure and finger holes (or the bore end) as illustrated in Fig. 2. Of course, the conditions for sound production should have been satisfied [9]. If so, the air jet operates as a growing wave affected by the bore resonance [10]. Although the essential importance of the edge is well understood by

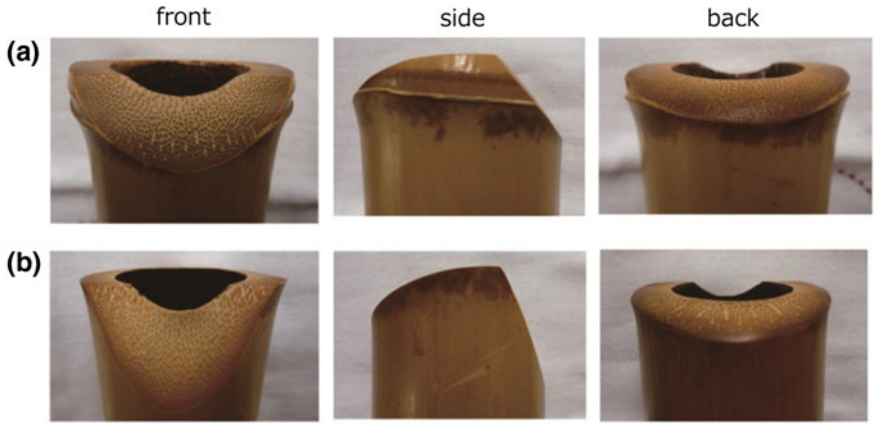


Fig. 1 Two examples of the shakuhachi embouchure edge. **a** A shallow edge with a short cut; **b** a deep edge with a long cut. Also, the construction of the back side is different between them

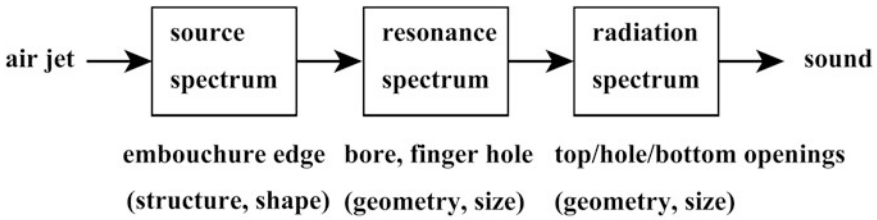


Fig. 2 Harmonic generation in the shakuhachi. The source spectrum formed by the interaction between the air jet and the edge is modified by resonance and radiation characteristics of the bore and finger holes

the player and the maker, scientific research on the flow acoustics around the real shakuhachi edge is still a future work [11].

Tone holes: In contrast with modern Western instruments with many tone holes (e.g. the clarinet, oboe, and flute has 24, 23, and 13 tone holes, respectively), the shakuhachi has only five tone holes traditionally. This means a decisive importance of cross (or fork) fingerings in the playing of it. A Japanese physicist, Torahiko Terada (1878–1935), first carried out an accurate measurement of its intonation [12]. He carefully measured pitch frequencies in the first and second registers for 32 fingerings, and directed attention to the octave balance.

If his intonation table is extensively examined, it is known that there are many cases where cross fingerings cause pitch sharpening instead of usual pitch flattening. Because the pitch sharpening due to cross fingerings is the reverse of conventional pitch flattening [9, 13–15], it may be called an *intonation anomaly* [16]. The acoustics of this intonation anomaly will be described later.

Inner bore: Yoshinori Ando (1928–2013) actively and accurately investigated the interrelation between the shakuhachi bore geometry and the resulting tones. He measured and calculated the input admittance of normal fingerings based on X-ray photography of the inner bore [17–20]. According to his research, there are four fundamental types of the bore geometry (the inner radius distribution along the bore) and major differences between classical (komuso) and modern shakuhachis.

Modern shakuhachis are often used in ensemble music and their exact tuning is required. As a result, diaphragms inside a bamboo pipe are completely removed, and then the inner pipe wall is shaved a little and pasted with a kind of clay consisting of polishing powder, *urushi* (Japanese lacquer), and water. The pasted surface in dried and solid condition is carefully polished up. This series of works may be called *ground-paste finish*. Also, the culm is divided between the third (counted from the bottom) and fourth tone holes in advance for the convenience of this ground-paste finish. Thus it has become easy to adjust the inner bore geometry, whose acoustical effects will be described later.

On the other hand, the original construction method of the shakuhachi had no ground-paste finish applied. The diaphragms were not completely removed and small ridges were retained on the inside nodes [21, 22]. These remaining portions of the diaphragms subtly affect the intonation and produce natural tones, which cannot be heard in modern ground-pasted shakuhachis. Most classical shakuhachis are *ground-paste free*.

It should be also noted that the bore is not divided and finger holes are undercut in classical shakuhachis. The length of the shakuhachi varies, although the standard length is 54.5 cm (pitched in D_4) as mentioned above. Recently, a ground-paste-free shakuhachi longer than *two-shaku and five-sun* (75.8 cm, $A^b{}_3$) has been preferred for personal deeper introspection.

2.3 Sound Examples

A few sound examples from a ground-paste-free shakuhachi [length: two-shaku and three-sun (69.5 cm); top bore diameter: 28 mm; bottom bore diameter: 23 mm; pitch: $B^{b}{}_{3}$] are shown in Fig. 3. This shakuhachi was made by Johzan Iso who made the one whose bore geometry was depicted in Fig. 4b. Its bore geometry is similar to Fig. 4b except for wider and smoother finish near the bottom. Also, its total view and the edge structure are given in Fig. 1.6 of Ref. [22] and Fig. 1a, respectively. All of these sound examples are played in the first register by the

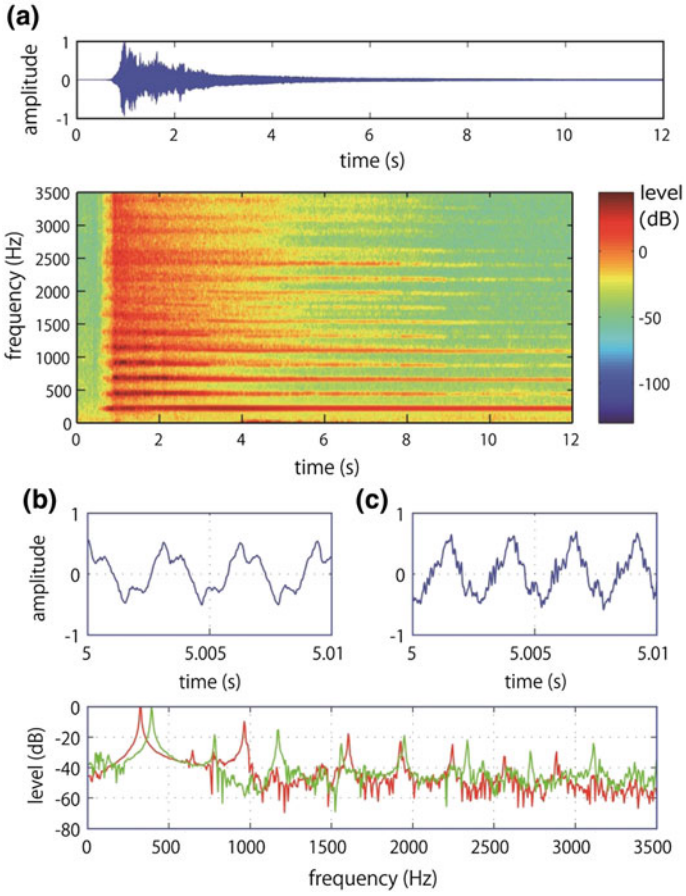


Fig. 3 Tone examples of the two-shaku three sun shakuhachi. **a** *Ro* (all tone holes closed) blown with *muraiki* during the starting four seconds; **b** *Wu* (the first and third tone holes opened half) with *meri* blowing; **c** *Ri* (the third and fourth tone holes opened) with normal blowing. The corresponding spectrum of the *Wu* and *Ri* tones is shown in the bottom frame by the *red* and *green lines*, respectively

author. Fingerings of (a) all holes closed, (b) the first and third holes opened half, and (c) the third and fourth holes opened completely are used in Figs. 3a–c, respectively.

Generally called *muraiki* (a rough and strong blow) is applied to the all-hole-closed fingering (called *ro*) during about four seconds from the starting transient in Fig. 3a. The spectrogram of the lower frame indicates the temporal change of the relative strength (in dB) of tonal components by color. The fundamental frequency varies from 224 Hz at the transient to 219 Hz of *pianissimo* playing near 10 s, and the pitch is closer to A_3 rather than B^b_3. This is due to a thick bore and the player's blowing way. The muraiki brings about very strong harmonics (from the second to the fifth) and a few inharmonic spectra above the seventh harmonic. Also, strong wind noise is involved from above 1.3 kHz to about 2.7 kHz. After the blowing becomes normal, even harmonics are very weak and odd harmonics (the fundamental, the third, and the fifth) are predominant. The harmonic structure with stronger odd harmonics is a distinguished character of the classical ground-paste-free (komuso) shakuhachi [17, 20].

The *meri* (or down) blowing given by pulling down the jaw is applied to the fingering *wu* (the first and third tone holes are half opened) in Fig. 3b. The fundamental frequency of a steady tone is 319 Hz, and the pitch is close to E^b_4. As shown in the spectrum diagram of the lower frame, this tone (drawn by the red line) almost lacks the second and fourth harmonics. The normal blowing is applied to the fingering *ri* (the third and fourth tone holes are opened) in Fig. 3c. The fundamental frequency of a steady tone is 389 Hz, and the pitch is close to G_4. This tone (drawn by the green line in the spectrum diagram) contains rich harmonics, while the third and fifth harmonics are slightly predominant. Tone *wu* brings a blue, melancholic feeling; tone *ri* a cheerful, fine feeling.

2.4 *Acoustical Differences Between Classical and Modern Shakuhachis*

Ando [19] investigated bore geometries and dimensions of about 70 shakuhachis and classified them into four types. Furthermore, he intensively measured and calculated input admittances (i.e., resonance characteristics) of six shakuhachis typical of four types [17, 18]. Essential results of his research are summarized below.

Bore shape patterns observed in modern and classical shakuhachis are depicted in Fig. 4a, b, respectively. These were classified as "type 1" and "type 4" by Ando [17, 18], respectively. The "type 2" is a significant enlargement near the bottom of "type 1"; the "type 3" seems to be a relaxation of "type 4" around the bamboo nodes. Major differences between Figs. 4a, b are (1) small/large bore diameter, (2) convergent/divergent bore from the embouchure (the 1st node) to the 2nd node (located near 190 mm from the edge), and (3) without/with abrupt changes at the

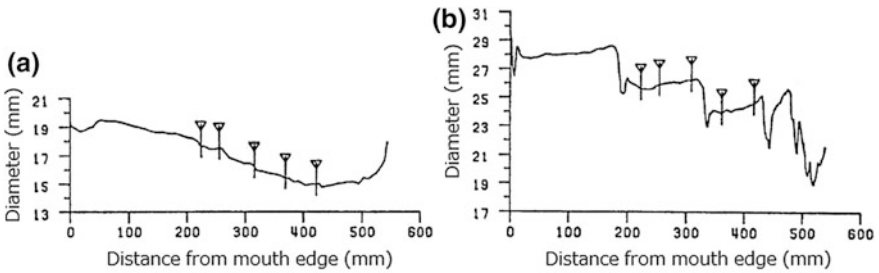


Fig. 4 Inner bore shape patterns of a modern shakuhachi (**a**) and a classical (*komuso*) shakuhachi (**b**) [17, 19]. Also, five tone-hole positions are indicated by the *vertical line*

nodes. Roughly speaking, modern shakuhachi of "type 1" has continuous convergent bore like the recorder and the baroque flute except for the portion near the bottom, while classical shakuhachi of "type 4" has an distinctive bore shape consisting of a few cylindrical pipes with stepwise decrease in diameter. Some important comments are given by Simura [21] from his long research experience.

The calculated input admittances of these modern and classical shakuhachis are shown in Fig. 5, respectively. Cases of two common fingerings *chi* (the first to third tone holes are open; A_4) and *ri* (the third and fourth tone holes are open; C_5) are exemplified, although Ando [17] calculated for six basic fingerings. The bore shape

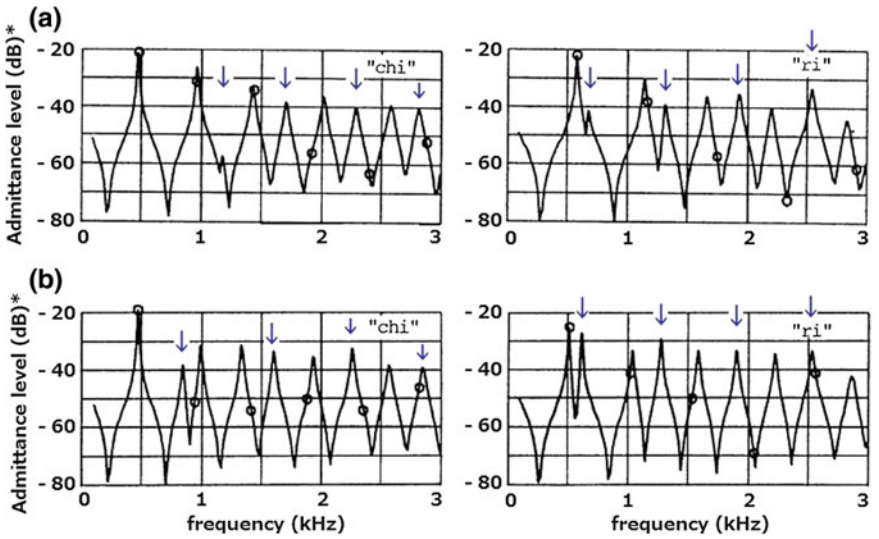


Fig. 5 The calculated input admittances of a modern shakuhachi (**a**) and a classical shakuhachi (**b**) [17]. Fingerings are basic ones, *chi* and *ri*. The symbol *open circle* indicates the harmonics of the fundamental frequency that is given by the first admittance peak. The *vertical arrows* suggest the upper-bore intermediate modes or the lower-bore modes. The symbol *asterisk* attached to "Admittance level" indicates the level relative to 1 SI unit (1 m^2 s/kg)

was approximated by many cylindrical segments in order to apply the transmission line theory [18, 23] and the lumped T circuit representations of the open/closed tone holes [18, 24, 25]. The numbers of cylindrical segments, which were determined based on the criteria of the calculation precision [18], were 67 and 168 of the above modern and classical shakuhachis, respectively [17].

The input admittance curves of Fig. 5 are rather complicated. This is probably due to the contribution of the lower bore below the top open tone hole (cf. next subsection). Although Ando [17, 18] suggested such a contribution, the present author would like to add more suggestive comments below based on the research of cross fingerings in the shakuhachi [16].

The fundamental frequency of basic fingerings is given by the first peak of the input admittance curve [17]. Harmonics are indicated by the symbol ○ marked on the curve in Fig. 5. As shown in Fig. 5a, the second and third harmonics of fingering *chi* are located near the tops of the second and third peaks, though the fourth, fifth, and sixth harmonics deviate from the curve peaks. It should be noted that another series of peaks appears between the peaks that give harmonics as indicated by the vertical arrow. These peak frequencies might be caused by the resonance of the intermediate mode of the upper bore (i.e. the pipe above the top open tone hole) (cf. f_{34} in Fig. 10a) or by the resonance of the lower bore (i.e. the pipe below the top open tone hole) (cf. f'_{2-} in Fig. 10a). It is confirmed that Fig. 5a agrees with Fig. 10a very well.

The third tone hole is the top open tone hole in the case of normal fingering *chi*. The length of the upper and lower bores is about 320 and 220 mm, respectively. Because the end corrections at the embouchure and the open top tone hole (roughly estimated as 30 mm in total) should be added, the fundamental frequency is calculated as 345,000/(350 × 2) = 493 Hz. Because the lower bore seems to generate no radiation, the end correction should be negligible, then its fundamental frequency is calculated as 345,000/(220 × 2) = 784 Hz if the first and second tone holes operate as the closed ones. Although this frequency is not observed in Fig. 5a, it might be observed in Fig. 5b. The difference between two figures around 1 kHz might suggest the difference in the acoustic coupling between the upper and lower bores occurring at the third open tone hole. Particularly, Fig. 5b on fingering *chi* suggests that the mutual repelling might be occurred between the second mode of the upper bore and the first mode of the lower bore because of the raised frequency of the second mode of the upper bore in comparison with Fig. 5a. However, a very small peak near 1.2 kHz in Fig. 5a should be the lower second mode of the lower bore (see f'_{2-} in Fig. 10a).

Such a modal repelling may be seen between the first modes of the upper and lower bores in Fig. 5b on fingering *ri* whose top open tone hole is the forth. The length of the upper and lower bores is 260 and 280 mm, respectively. Assuming the end corrections, the fundamental frequencies are 345,000/(290 × 2) = 594 Hz and 345,000/(280 × 2) = 616 Hz, respectively. The fundamental frequency of the classical shakuhachi might be reduced a little by the modal repelling. Since fingering *ri* gives tone holes closed below the third one, the effect of the lower bore on the admittance curve is more significant.

The admittance curve of the shakuhachi (exactly its upper bore) is apparently *inharmonic* due to (1) the bore-shape perturbation and (2) frequency characteristics of energy dissipation along the wall boundary and energy radiation from the open ends. This inharmonic series of the peaks determines the harmonic content of the tone generated if the effect of acoustic inertance lumped at the embouchure end can be considered properly [8, 17, 26].

According to Ando [20, 27], a significant tonal difference between modern and classical shakuhachis can be expressed as $L_e - L_o$, where L_e denotes the averaged level of even (2nd, 4th, and 6th) harmonics and L_o the averaged level of odd (3rd, 5th, and 7th) harmonics. For basic six fingerings, the classical shakuhachi indicates the dominance of odd harmonics and gives around null $L_e - L_o$ values, while the modern shakuhachi indicates the dominance of even harmonics and gives apparently positive $L_e - L_o$ values. Sound examples *ro* (after 8 s) and *ri* shown in Fig. 3a, c give an apparently negative $L_e - L_o$ value and a slightly negative $L_e - L_o$ value, respectively. The down blowing strongly emphasizes this tendency by almost removing even harmonics as shown in Fig. 3b for the cross fingering *wu*. Moreover he [20] demonstrated that the difference in the $L_e - L_o$ value substantially depends on the bore shape from the embouchure end to a point 110 mm down. A generally decreasing diameter (a convergent bore) as shown in Fig. 4a yields apparently positive $L_e - L_o$ values; a generally increasing diameter (a divergent bore) as shown in Fig. 4b yields around null $L_e - L_o$ values.

The effects of small ridges remaining on the inside nodes upon shakuhachi tones is a very interesting topic [12]. However, it is rather difficult to separate them from the effects of overall bore shape. The effects of small ridges, which can be estimated by Rayleigh's perturbation theory [9, 28], seems to be insignificant compared with the effects of overall bore shape (cf. Fig. 4b). Anyway, this problem should be solved in the near future.

2.5 *Intonation Anomaly Due to Cross Fingerings*

A decisive importance of cross fingerings in the playing of the shakuhachi is easily understood from its only five tone holes. As briefly mentioned in Sect. 2.2, cross fingerings often cause pitch sharpening instead of usual pitch flattening. This *intonation anomaly* due to cross fingerings, which is observed in the recorder and Baroque flute too, usually appears in the second register. The acoustics of the intonation anomaly [16] is described below.

Terada [12] investigated tonal octave balance on 32 fingerings including 26 cross fingerings, and Yoshikawa and Kajiwara [16] intensively studied 7 fingerings including 5 cross fingerings on the basis of the pressure standing wave along the bore and the input admittance. It is important to identify and discriminate the input-admittance spectra between the upper and lower bores from the standing-wave patterns. In this subsection, the results on three fingerings (*chi*, *wu*, and *wu3*) whose top open tone-hole is the third one (see Fig. 6) are illustrated. Also,

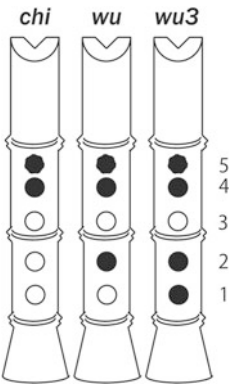


Fig. 6 Three fingerings of the shakuhachi treated in this section

the bore shape of a standard shakuhachi used for the experiment and numerical calculation is depicted in Fig. 7.

At first, the playing experiment is effective. The playing frequencies of each fingering are easily measured. In the first register three fingerings (*chi*, *wu*, and *wu3*) gave 444 Hz (A_4), 433 Hz (A^b_4), and 426 Hz (A^b_4), respectively (room temperature was about 23 °C). In the second register the three fingerings gave 898 Hz (A_5), 853 Hz (A^b_5), and 920 Hz ($A^\#_5$), respectively. Furthermore, in the third register the three fingerings gave 1322 Hz (E_6), 1475 Hz (G^b_6), and 1472 Hz (G^b_6) [plus 1273 Hz (E^b_6)], respectively [16]. The underlined frequencies denote intonation anomalies. Other examples from different fingerings are shown in Ref. [16].

Secondly, the external blowing experiment is effective, too. The measurement of the internal pressure distributions (standing-wave patterns) can be carried out by moving a probe microphone (Brüel and Kjær type 4182) with a long probe tube (e.g. 570 mm in length and 1.25 mm in inner diameter) when the external drive is successfully done by using an exponential horn attached in front of the loudspeaker diaphragm (see Fig. 8). Resonance frequencies of a fingering are measured prior to

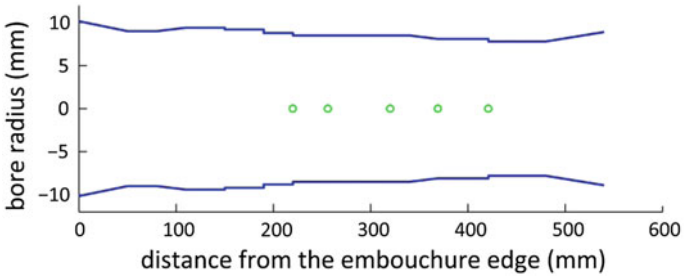


Fig. 7 Bore geometry of a modern shakuhachi treated in this section. The tone-hole positions are also indicated by the *circle*. The bore is approximated by ten cylindrical, two divergent conical, and two convergent conical tubes for numerical calculation

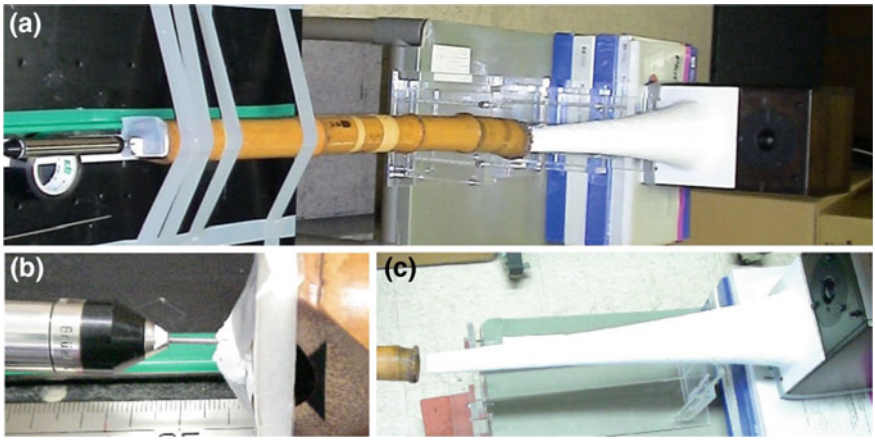


Fig. 8 Setup of the blowing experiment for measuring the pressure standing waves along the air column of the shakuhachi. **a** Total view; **b** close-up of a probe microphone and the embouchure; **c** close-up of the shakuhachi bottom and an exponential horn whose shape was designed to have the cutoff frequency at about 200 Hz

the standing-wave measurement. The details of measurement method and result are given in Ref. [16].

Thirdly, the calculation of the input admittance is also very effective as mentioned in Sect. 2.4. The conventional transmission matrix (*T*-matrix) method has been applied to the bores of woodwinds [17, 18, 23] and brasses [29–31]. Also, see the fifth chapter on the "acoustical modeling of mutes for brass instruments" involved in this book for the *T*-matrix formulation. On the other hand, the tone or finger hole is not simple as the bore, and we have a long and extensive history on acoustical tone-hole research [9, 14, 24, 25, 32, 33]. In this section and Ref. [16] new results given by Lefebvre and Scavone [33] are applied to the input-impedance calculation. The tone-hole position is indicated in Fig. 7, the tone-hole diameter is about 10 mm, and the tone-hole length is about 7.5 mm [16]. Moreover, the calculation of the internal pressure distribution along the bore can be carried out on the basis of the *T*-matrix formulation with the tone-hole matrix representation. This internal pressure calculation was first explicitly formulated by Ebihara and Yoshikawa [31] on brass instruments.

The results of the external driving experiment and the numerical calculation based on the *T*-matrix method are shown in Fig. 9. Fingerings *chi*, *wu*, and *wu3* are used. Note that the calculation is done at the same frequency as the measured one by adjusting the embouchure end correction except for f_4 (1903 Hz) and f_4 (1880 Hz) in Figs. 9a, c, respectively. The upper-bore modes are illustrated, where the upper-bore mode is usually defined as the standing wave indicating larger amplitude in the upper bore and satisfying the resonance conditions (the pressure minima) at both ends of the upper bore above the third tone hole. Also, $f_n \approx nf_1$ for

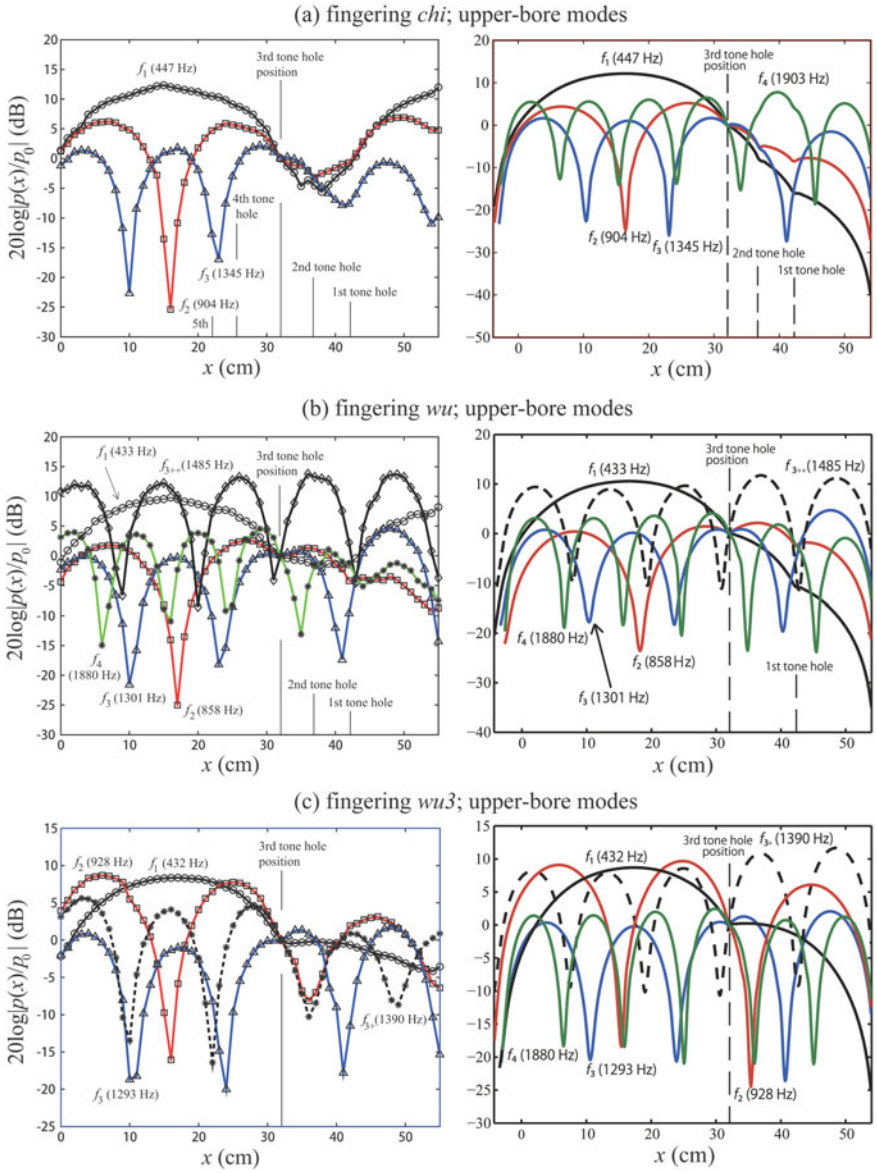


Fig. 9 Results of the measurement (*left column*) and numerical calculation (*right column*) on the internal standing-wave patterns [16, 34]. The distributions of the upper-bore modes are depicted. **a** Normal fingering *chi* (the first to third tone holes are open); **b** cross fingering *wu* (the first and third tone holes are open); **c** cross fingering *wu3* (only the third tone hole is open). Note that the acoustic pressure $p(x)$ is normalized by the pressure p_0 at the third tone hole

the mode order n = 1, 2, 3, and 4. Subscripts such as "+" and "++" are used to discriminate multiple modes in the same mode, such as f_3 and f_{3+} ($f_{3+} > f_3$).

However, there are a few exceptions: (1) the f_{3++} mode (1485 Hz) in fingering *wu*, (2) the f_3 mode (1293 Hz) in fingering *wu3*, and (3) the f_{3+} (1390 Hz) in fingering *wu3*. The first two modes do not satisfy the resonance condition at the open third tone hole, but they satisfy the resonance condition at the bore bottom. Therefore, they should be regarded as the *whole-bore mode* instead of the upper-bore or lower-bore mode. It should be noted that these two modes are actually played as tones with frequencies 1475 Hz (G^b_6) and 1273 Hz (E^b_6) respectively as mentioned above. Although the third one f_{3+} (1390 Hz) satisfies the resonance condition at the open third tone hole in the blowing experiment, it does not satisfy the resonance condition at the bore bottom. It seems that such a mode can be measured due to the external drive near the bore bottom. On the other hand, numerical calculation indicates that this f_{3+} (1390 Hz) violates the resonance condition at the third tone hole if it is considered as the upper-bore mode. However, if it is considered as the lower-bore mode, it satisfies the resonance conditions both at the bore bottom and the third tone hole. Only this f_{3+} (1390 Hz) mode brings about the major discrepancy between the experiment and the calculation. Except this mode and the distributions along the lower bore, the agreement between the experimental and calculated results shown in Fig. 9 is very high.

The calculated result of the input admittance $|Y_{IN}|$ is given in Fig. 10. It should be noted that small peaks f'_{2-} and f'_1 appear in Figs. 10a, c, respectively. These two peaks with the prime possibly indicate the lower-bore modes. Because the third tone hole is located at 220 mm from the bore bottom, the upper-bore physical length is 320 mm. Then, the first mode of the upper and lower bores is given as f_1 (432 Hz) and f'_1 (681 Hz), respectively (see Fig. 10c). Since the f'_{2-} seems to be quite lower than the assumed second mode of the lower bore, subscript "−" is added in Fig. 10a.

Although the f_2 (837 Hz) in cross fingering *wu* is lower than the f_2 (903 Hz) in normal fingering *chi*, the f_2 (945 Hz) in cross fingering *wu3* is appreciably higher than that in normal fingering. Thus the f_2 (945 Hz) indicates the intonation anomaly. Also, the f_{3++} (1473 Hz) and f_{3++} (1494 Hz) in Figs. 10b, c may be regarded as the intonation anomaly compared with f_3 (1318 Hz) in Fig. 10a.

In order to demonstrate the intonation anomaly in clearer fashion, the calculated standing-wave patterns for three fingerings in Fig. 9 are re-drawn for the respective mode in Fig. 11 [16, 34]. Figure 11a is on the first mode, where the pressure along the lower bore below the open third tone hole becomes higher as the second and first tone holes are closed in succession in fingering *wu* and *wu3*. Also, a weak kink of the pressure amplitude, which indicates the phase change due to the partial reflection, is seen at the open tone hole. These patterns well illustrate the typical (or conventional) effect of cross fingerings, which yields the descent of the resonance frequency.

On the other hand, cross fingering *wu3* produces a very deep trough near the closed second tone hole, as shown in Fig. 11b for the second mode. Also, the kink at the open third tone hole is inappreciable for this second mode. As a result, the

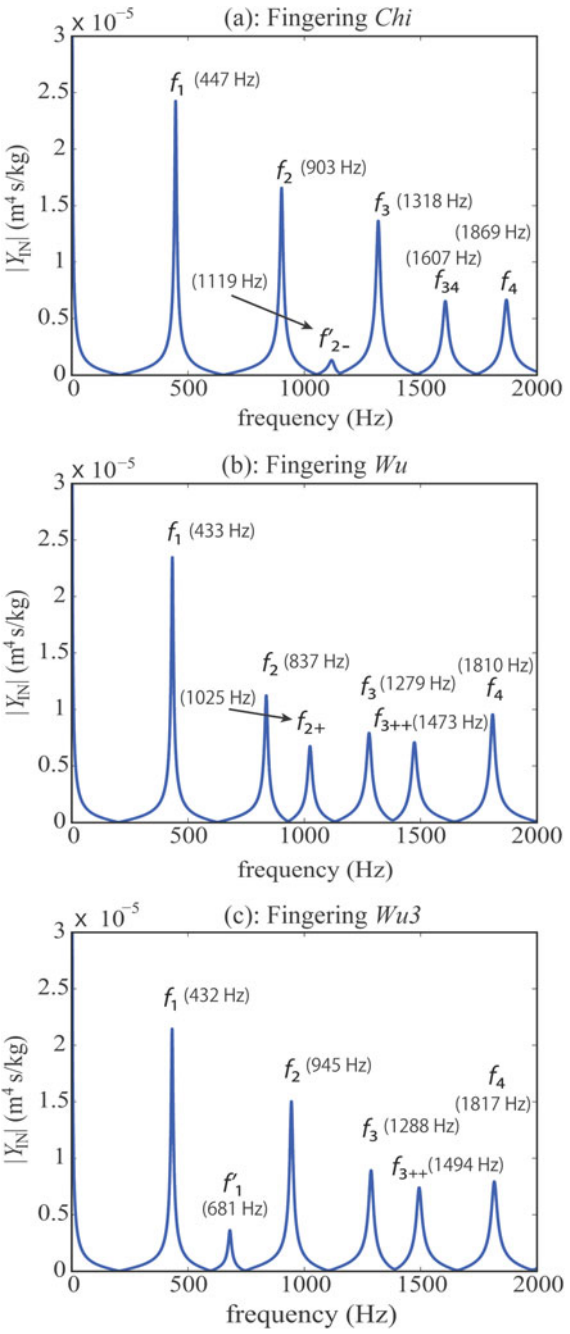


Fig. 10 The calculated absolute value of the input admittance $|Y_{IN}|$ for three fingerings [16]. **a** Normal fingering *chi*; **b** cross fingering *wu*; **c** cross fingering *wu3*

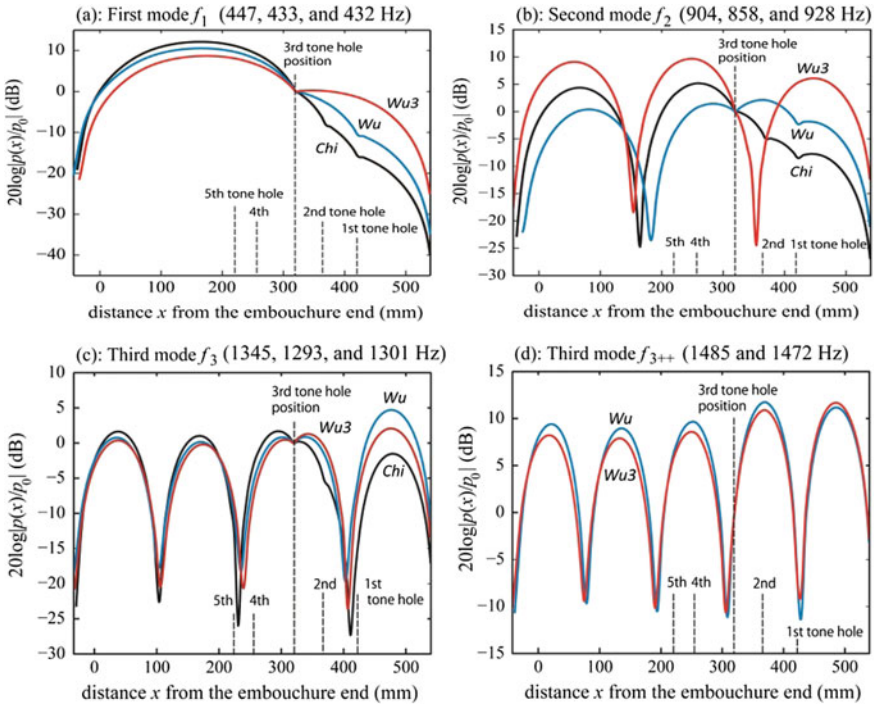


Fig. 11 Standing-wave patterns for the respective mode given by the three fingerings [16, 34]. The mode frequencies noted in each frame are in order of fingering *chi*, *wu*, and *wu3*

wavelength of this mode by *wu3* is significantly shorter than those by *chi* and *wu*. At this time, the clear third mode is formed along the whole bore and the intonation anomaly is induced. It may then be understood that the lower bore is almost completely coupled with the upper bore instead of being separated at the top open tone hole. The whole-bore mode is thus formed.

Although each third mode f_3 seemingly forms the fourth mode along the whole bore as shown in Fig. 11c, the kink (phase change) at the open top tone hole is stronger than that of the first mode shown in Fig. 11a. Then, the complete coupling at the top open tone hole is obstructed, and the intonation anomaly does not occur, as noted in the measurement and playing results.

However, cross fingerings *wu* and *wu3* easily yield the higher third mode f_{3++}, as shown in Fig. 11d. It should be noted that this higher third mode was really played by the player. Therefore, this mode may be regarded as an upper-bore mode, but it violates the resonance condition at the top open tone hole for the upper-bore mode. Moreover, the pressure amplitude along the lower bore is larger than that along the upper bore. Hence, this f_{3++} might be a lower-bore mode. In either case, it is essential that the whole-bore mode (the fifth mode) due to the complete coupling between the upper and lower bores is formed through the continuity (no phase change) at the top open tone hole and then the intonation anomaly is induced.

The intonation anomaly is derived from the complete coupling between the upper and lower bores through an open top tone hole. At this time, the discrimination of the upper-bore mode from the lower-bore mode is rather difficult as indicated in Fig. 11d. This may be because the third mode frequency of the upper-bore resonance is very close to the second mode frequency of the lower-bore resonance. In general, the intonation anomaly may be deduced when one of the resonance frequencies of the upper bore (from the embouchure end to the outer end of the top open tone hole) is very close to one of the resonance frequencies of the lower bore (from the bore bottom to the outer end of the top open tone hole). This strongly depends on the position of the top open tone hole [16]. Under such a situation, the modal interaction or mutual repelling (supposed in the explanation of Fig. 5b) in a coupled resonance system might be occurred.

Also, since the top open tone hole functions like a closed tone hole (cf. Fig. 11b, d) when the intonation anomaly occurs, the cutoff frequency of the open-tone-hole lattice [9, 15, 32] might be involved. The calculated cutoff frequency was about 1270 Hz when averaged geometrical values on the bore and tone holes are applied [16]. The modes penetrating into the lower bore such as f_3, f_{3+}, and f_{3++} might be related with the cutoff frequency. More detailed discussion on the physical mechanism causing the intonation anomaly and its modeling leading to our adequate understanding will be an important issue from the viewpoint of musical acoustics.

3 The Nohkan

The transverse bamboo flute, *nohkan* with seven finger holes is usually performed in ensemble with two-head drums (larger one is called *ohtsuzumi*; smaller one *kotsuzumi*) in Japanese traditional musical drama, *noh*. Its unique acoustical properties are described in this section.

3.1 Brief History

Four transverse flutes have been preserved in the *Shosoin* warehouse. They have seven tone holes, while the transverse flute *dizi* in China and *taegum* in Korea have six tone holes. Some varieties of the Indian flute *bansuri*, which dates in India from the first century AD at the latest, have six or seven tone holes [35]. According to Hayashi [4], flutes with seven tone holes were played in the secular music during the *Han* dynasty (206 BC–220 AD) of China, and they were introduced to Japan. Although the transverse flute first appeared in the *Han* period in China, its origin could be found in India [4, 36].

These Shosoin flutes were linked with the *ryuteki* played in the court music (*gagaku*), and furthermore were brought into the nohkan. Moreover, similar flutes were propagated from the Korean peninsula in the middle age. At last, the form of Japanese transverse flutes was decisively fixed [36]. Nevertheless, the origin of

Japanese transverse flutes (*fue*) still has many mysterious aspects [36]. A close relation between the nohkan altissimo tone (called *hishigi*) and the iwabue tone was mentioned in Introduction. The role of the iwabue seemed to lay down the god or ghost to the earth or this world. Most of noh dramas almost always have stories connecting this world with that world. The role of the nohkan seems to let the audience feel a premonition of this connection through its very high pitch and very solid timbre. On the other hand, there is another opinion that the nohkan altissimo tone is just only a signal to give the main players the timing for the entrance onto the stage [3, 36].

3.2 *Unique Structural Properties*

The external views of a nohkan are shown in Fig. 12. The top and side views are given in Fig. 12a, b, respectively. The total length of the nohkan is 410 mm, but the bore for the resonance is only 312 mm long as shown in Fig. 12b. The tube from the left edge of the embouchure (or mouth) hole to the left closed end is filled with bees-wax in which lead or iron bar is embedded. The bore diameter at the left edge of the embouchure hole is about 18 mm. However, it should be noted that the nohkan is different from one to the other. This is because the nohkan has no definite tuning pitch and it never be played with other nohkans in ensemble.

The embouchure hole is oval and large (the diameter is 19 mm long and 16 mm wide) as shown in Fig. 12c. Its edge against which the player's breath is blown is rather shallow so that the player can apply very strong blowing pressure in the second and third registers [37, 38]. The resulting forceful attack generates a great deal of wind noise. This is definitely different from Western flutes. Also, finger holes are not flat but curved against the external surface of the instrument as shown in Fig. 12b. This is because the tone holes of the nohkan (and the ryuteki) are covered by the middle joints of the fingers and because the half-hole and partly-raising-finger techniques for subtle pitch and timbre adjustments [3, 39] are often used. The diameter of tone holes is 12–13 mm.

Although the nohkan resembles the ryuteki, the nohkan has a thin tube inserted between the embouchure hole and the first (counted from the top) tone hole as illustrated in Fig. 12d. This inserted tube is called *nodo* (throat), which constricts the bore and makes the notes of the second octave increasingly flat to the lower octave as the scale is ascent [27, 35, 37, 38]. In other words, there is no concept of "octave" in playing the nohkan. The origin of this unique device is still obscure. Possibly one of nohkan makers found incidentally its interesting effects on the pitch and timbre during the repairing process [36]. Anyway, this nodo should be the Japanese invention probably carried out in the *Muromachi* period (1334–1573) during which the noh play was largely developed. Note that a tube section around the nodo is cut in half near the middle of the nodo so that the nodo is smoothly inserted. The narrowed bore shape of the nodo is approximately depicted by the dashed line in Fig. 12d, but there are some different bore shapes, one of which is almost straight except for both ends.

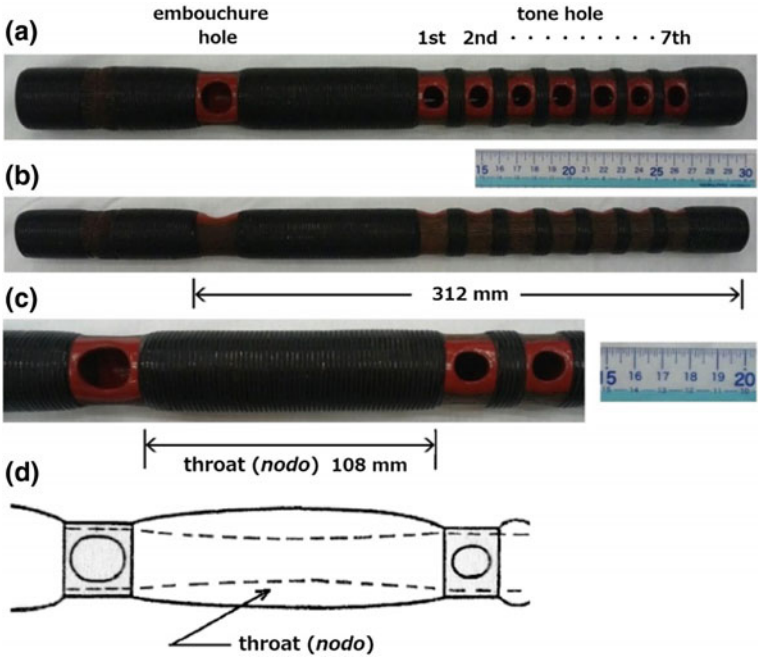


Fig. 12 External views of a nohkan (offered by professional player DenshoTosha). **a** Top view; **b** side view; **c** close-up from the embouchure hole to the second tone hole; **d** illustration of the throat (*nodo*) inserted between the embouchure hole and the first tone hole. This illustration is based on Fig. 3 in Ref. [37]. This nohkan is quite longer than the averaged one

Especially on this flute, a note is characteristically attacked from well below, strongly and breathily [35]. Malm [3] described the uniqueness of the nohkan: "the indefinite quality of its tone and its music are eminently suitable for supporting the drama without interfering with the declamation of the poetry". This declamation is called *ji-utai* in Japanese and it is completely different from the Western singing but rather close to narrating or recitative chanting. Since the nohkan has no consistent pitch and no octave balance, there can be no deliberate relation between the pitches of the instrument and those of the vocal line [3]. On the basis of essentially indefinite characteristics of the nohkan and the ji-utai, the noh is played freely between this world and that world [40].

3.3 Sound Examples

Some tonal examples are given in Fig. 13. These tones were played by professional player Densho Tosha and recorded in a large studio of the Osaka University of Arts by Prof. S. Simura. Tones in the first and second registers when all tone holes are closed are shown in Fig. 13a, b, respectively. The upper frame shows the temporal

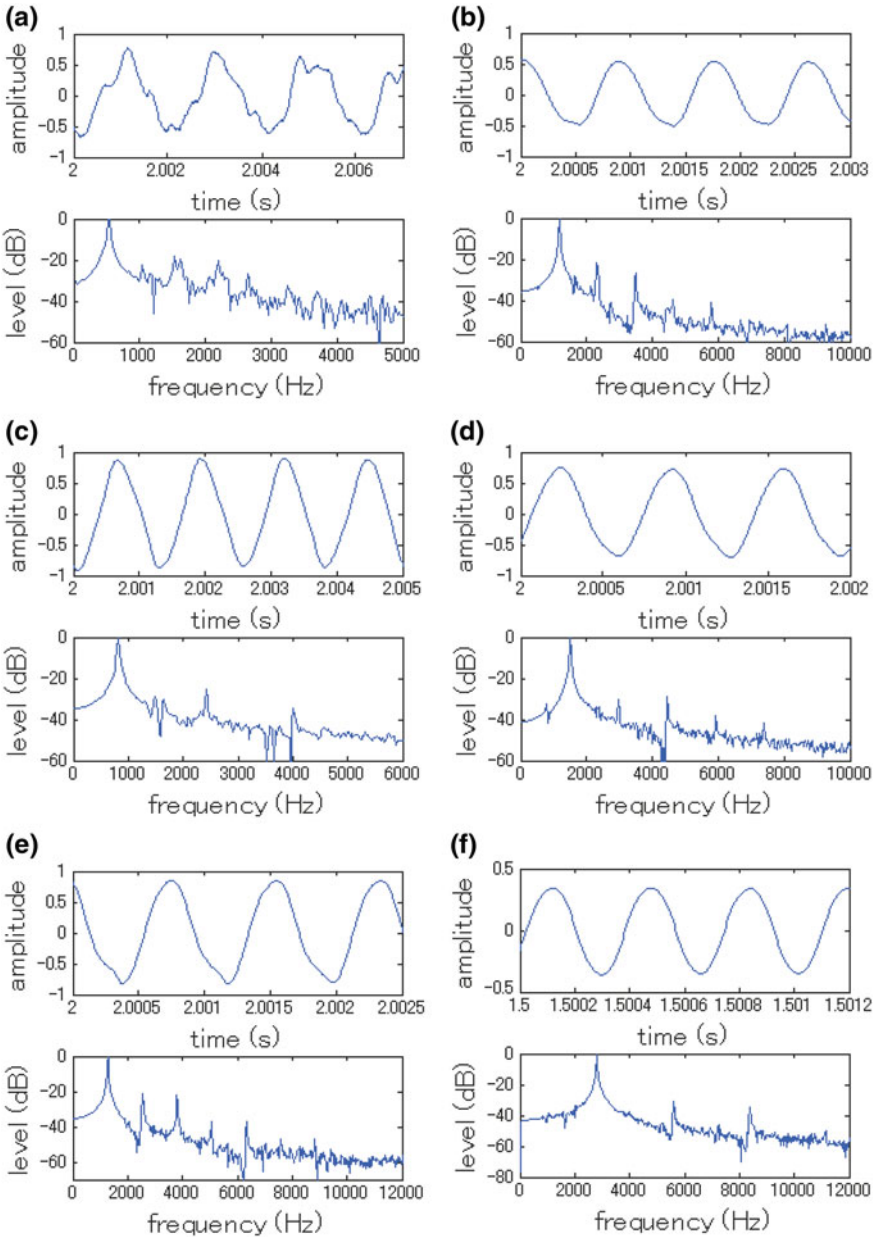


Fig. 13 Examples of steady-state tones by the nohkan. The spectral level is relative (normalized by the pressure amplitude of the fundamental). The details are given in the text

waveform and the lower frame the corresponding frequency spectrum. The fundamental frequency of the first register is 503 Hz (a little higher than B_4), and the overblown frequency in the second register is 1157 Hz (a little lower than D_6). This clearly indicates the octave enlargement. Also, the spectrum of the first-register tone shows indefinite harmonic structure. The lower harmonics are seemingly masked by wind noise from about 1–2.5 kHz. On the other hand, the second to fifth harmonics can be recognized in the second-register tone.

A cross fingering that opens the fourth to sixth tone holes produces the first-register and second-register tones shown in Fig. 13c, d, respectively. The respective tonal pitch is 794 Hz (G_5) and 1472 Hz ($G^b{}_6$). The octave is considerably shrunk. The second and fourth harmonics of the first-register tone cannot be clearly detected. Although the lower harmonics are observed in the second-register tone, the second harmonic is quite weak. Also, a small peak at 792 Hz is recognized. This peak probably reflects the first-mode resonance of the fingering.

A peculiar cross fingering that opens two extreme (the top and the bottom) tone holes produces the first-register tone shown in Fig. 13e. The tonal pitch is 1257 Hz ($E^b{}_6$). Since the pitch of the second-register tone (not shown here) is 2327 Hz (D_7), the octave shrink occurs. The sounding frequency of Fig. 13e gives the half-wavelength 137 mm, which corresponds to the resonating bore length. Therefore, this tone reflects the resonance of the throat itself, which brings unexpectedly quite rich harmonics (up to the seventh harmonic). Also, it can be understood that the total end corrections at both openings are estimated as about 29 and 40 mm for the first and second registers, respectively.

Another cross fingering that opens the second, fifth, and sixth tone holes produces the third-register tone as shown in Fig. 13f. This tone is called the *hishigi*, which means "to crush" or "to squash". This powerful piercing tone has an uncanny atmosphere. The tone frequency is 2775 Hz (F_7). The second and third harmonics are very weak, and the waveform is close to pure sinusoid. Also, we may observe small level changes from about 1 to about 2 kHz. These components might be related with the first and second modes of the resonance given by the fingering as noticed in Fig. 13d.

The octave balance between the first and second registers is depicted in Fig. 14, where the "octave" curve (connected by the square symbol) corresponds to twice the frequency of the first-register tone. The abscissa indicates the position of the top open tone hole given by the fingering. The octave enlargement occurs when all tone holes are closed (shown as "CL" on the abscissa) and when the top open tone hole is the seventh and the sixth. However, the octave shrink occurs when the top open tone hole is the fifth to the first. Like this, the throat makes the second-register tones increasingly flat as the scale is ascent and upsets the pure octave. As a result, the playing in the second register with subtle tone intervals becomes possible. The result of actually playing the nohkan shown in Fig. 14 well agrees with the measurement result of Ando [37, 38].

It is essentially important that the nohkan has been always pursued very solid and hard timbre. This tendency might be reverse the modern shakuhachi, while it

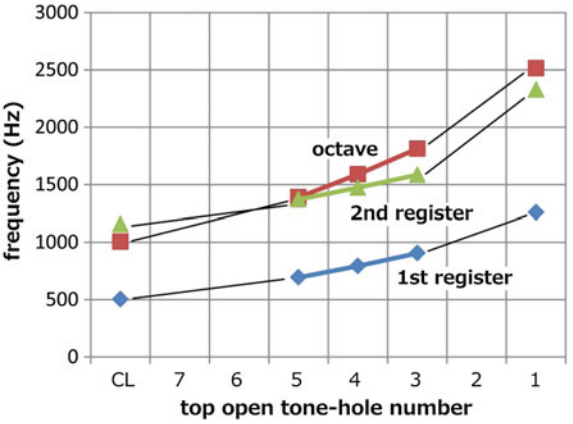


Fig. 14 Octave relation between the first and second registers based on the actually played tones. The "octave" curve gives the pure octave of the first-register tone

might be partially common to the old (*komuso*) shakuhachi, which however attaches importance to the first register. The inner wall of the nohkan is undercoated by the *urushi* lacquer 10–20 times, moreover the cinnabar urushi is coated over it 5–10 times [1]. This urushi coating seems to have deep relation with the hard timbre of the nohkan. The quality and the layer of the urushi lacquer seem to considerably affect the timbre of the nohkan and old shakuhachi [22].

3.4 Numerical Calculation on the Effects of Nodo

In order to understand the effects of the throat (*nodo*) more quantitatively, numerical calculation based on the *T*-matrix method was carried out on the throat shapes depicted in Fig. 15 [41]. A normal model of the throat, which is drawn by

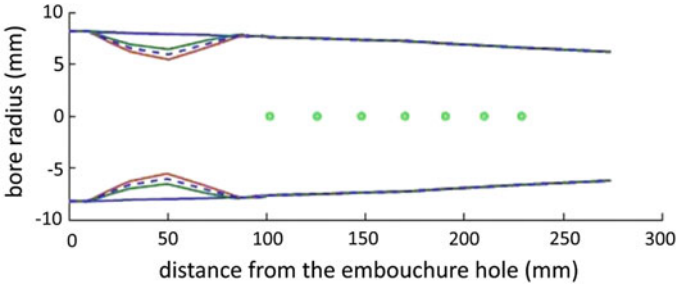


Fig. 15 Throat shape models for numerical calculation [41]. The *dashed line* is supposed to be a normal model seen in common nohkans. The *blue solid line* indicates no throat; the *green solid line* the throat widening the normal throat by 2 mm; the *red solid line* the throat narrowing the normal throat by 2 mm. The nohkan bore without the throat is convergent conical just as the piccolo. Also, the finger hole position is marked by the *green circle*

the dashed line, has a total length of 80 mm, and its center is located at 50 mm from the embouchure hole. If there is no throat, the overall bore (the blue line) is close to reverse (convergent) conical. The throat diameter at the center is varied from 14 mm (for a widened model depicted by the green line) to 10 mm (for a narrowed model depicted by the red line) through 12 mm (for the normal model).

At first the comparison between the nohkans with and without the throat (drawn by the dashed line and the blue solid line, respectively) is carried out. Four kinds of fingerings are used: fingerings A (all tone holes are closed), B (the fourth, fifth, and sixth tone holes are open), C (the first and seventh tone holes are open), and D (the second, fifth, and sixth tone holes are open). These fingerings were used in our playing experiment described in Sect. 3.3.

Results of numerical calculation are depicted in Figs. 16, 17, 18 and 19 for each fingering [41]. The dashed blue line is on the nohkan without the nodo (throat); the solid red line is on the nohkan with the nodo (throat).

The result of fingering A in Fig. 16a shows a definite frequency shift of the second mode resonance. The frequency raised by inserting the throat is 62 Hz. As a result, the octave enlargement occurs. It should be also noted that the reverse conical bore (without the throat) yields the octave enlargement (from 538 to 1100 Hz). This conical bore characteristic was applied to the Baroque flute to improve the octave shrink brought by the Renaissance cylindrical bore. Although Japanese transverse flutes have no headjoint, Western flutes always have it. Particularly its effect on the octave balance was well known historically. A good explanation of it is given by Benade [13] using his Fig. 22.11. The fact that the Japanese flutes have no such headjoints probably implies no concept of the octave in Japanese traditional flute music, particularly in the nohkan music.

The peak amplitudes of the first and third modes (f_1 and f_3) are larger than that of the second mode (f_2) when the throat is inserted. Also, the peak amplitudes of the modes above the fourth are quite low. These characteristics and the octave enlargement mentioned above of $|Y_{\mathrm{in}}|$ probably produce the tonal spectrum shown in Fig. 13a, although the third harmonic might be masked by the wind noise.

Interestingly enough, the half-wavelength at the embouchure-hole side is increased by inserting the throat as indicated in Fig. 16c. However, since the second resonance frequency is clearly ascent, the shorter half-wavelength at the bore-bottom side should be responsible for the pitch ascent. The inner pressure distributions of the first and third modes are not appreciably changed by inserting the throat except for the amplitude increase along the throat as shown in Fig. 16d, which causes the amplitude increase in the third mode of $|Y_{\mathrm{in}}|$.

On the other hand, fingering B gives the second mode which is not affected by the throat, but the first mode is increased by 38 Hz when the throat is inserted as shown in Fig. 17a. As a result, the octave shrink occurs. This result agrees with the playing experiment shown in Fig. 14. The peak amplitude of f_1 is decreased by the throat as suggested from the decrease in $p(x)$ shown in Fig. 17b. However, the peak amplitude of f_2 is significantly increased by the throat as suggested from the quite increase in $p(x)$ along the throat shown in Fig. 17c. The strength of the bore resonance seems to be affected by the throat from the results on fingerings A and B,

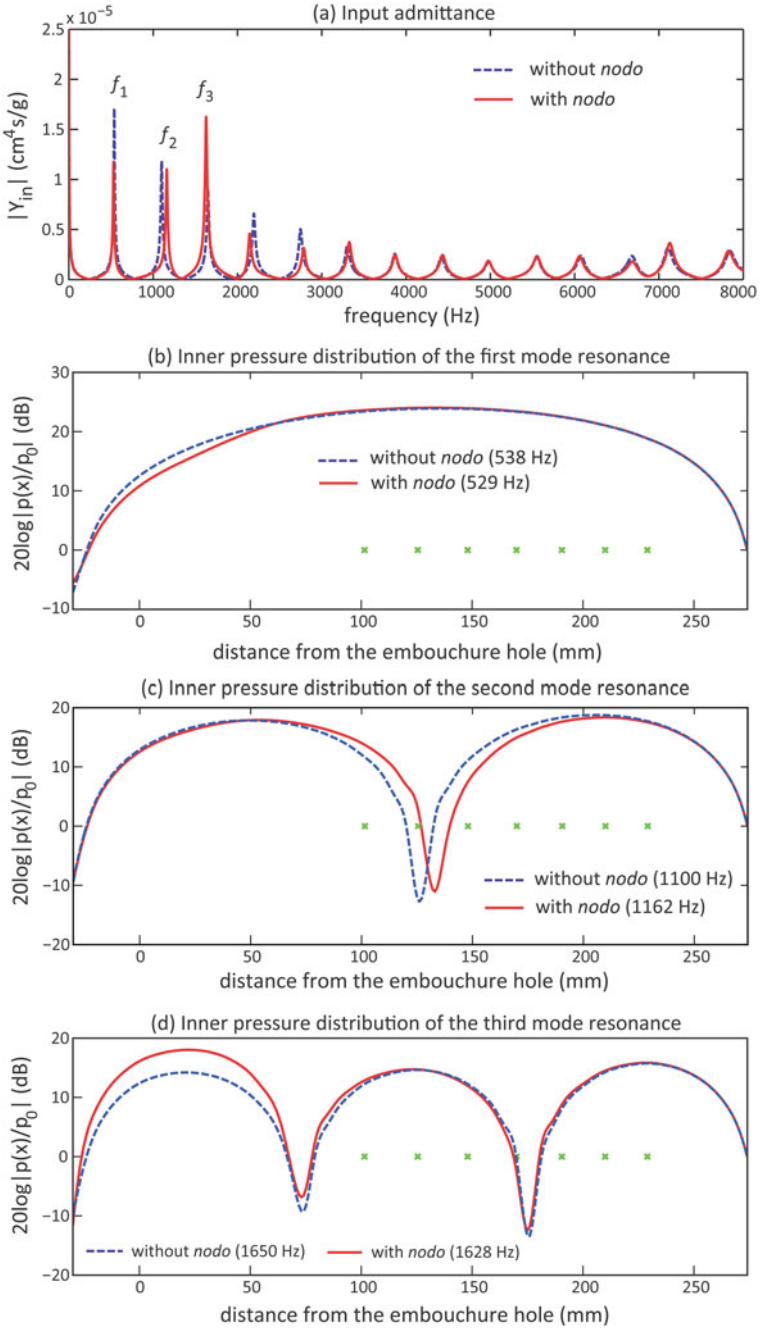


Fig. 16 Results of numerical calculation on fingering A [41]. **a** Input admittance; **b–d** inner pressure distribution along the bore in the first to third mode resonance, respectively. The pressure $p(x)$ at the position x from the right edge of the embouchure hole ($x = 0$) is normalized by that at the bore bottom. The symbol "×" indicates the closed tone hole

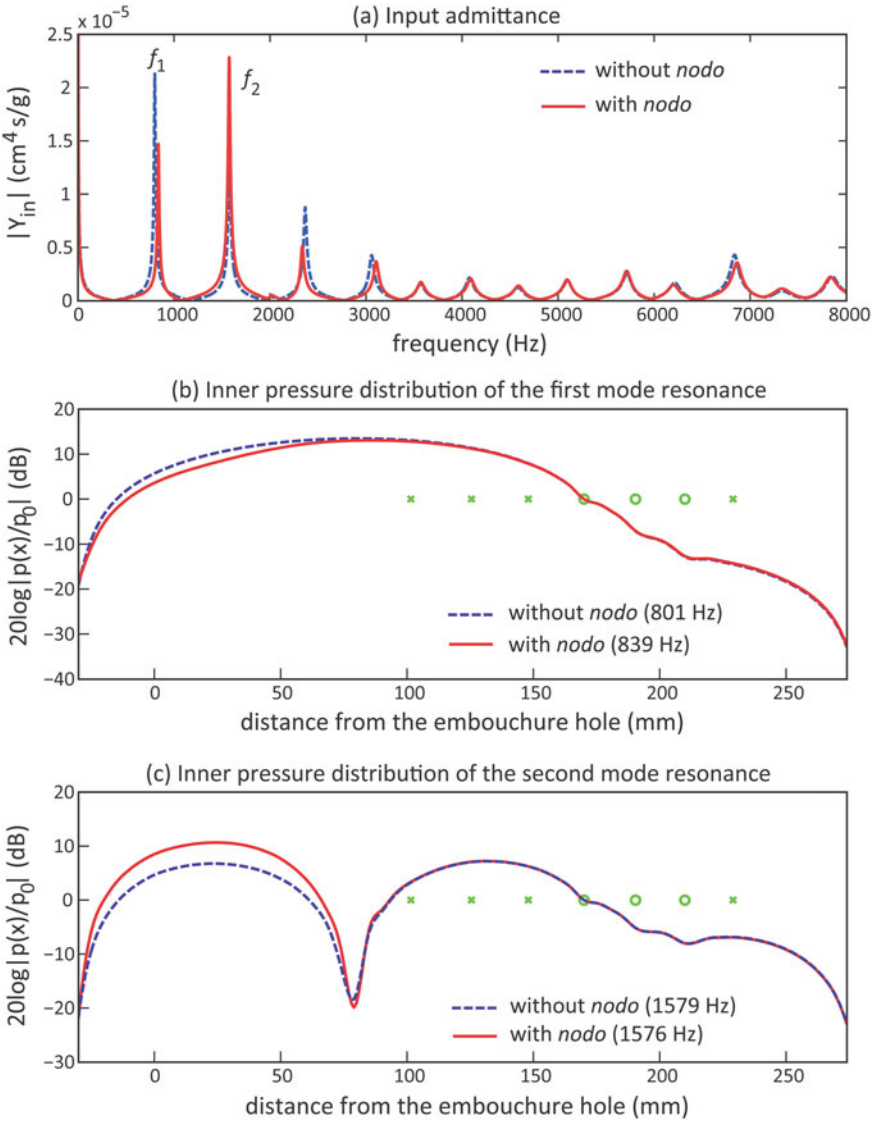


Fig. 17 Results of numerical calculation on fingering B [41]. **a** Input admittance; **b** and **c** inner pressure distributions along the bore in the first and second mode resonances, respectively. The pressure $p(x)$ is normalized by that at the top (fourth) open tone hole. The symbol "×" indicates the closed tone hole; the symbol "*open circle*" the open tone hole

particularly the frequency range of around 1600 Hz tends to be strengthened. The half-wavelength of 1600 Hz is 108 mm, which corresponds to the acoustical throat length with the embouchure-hole end correction. Also, it should be noted that the

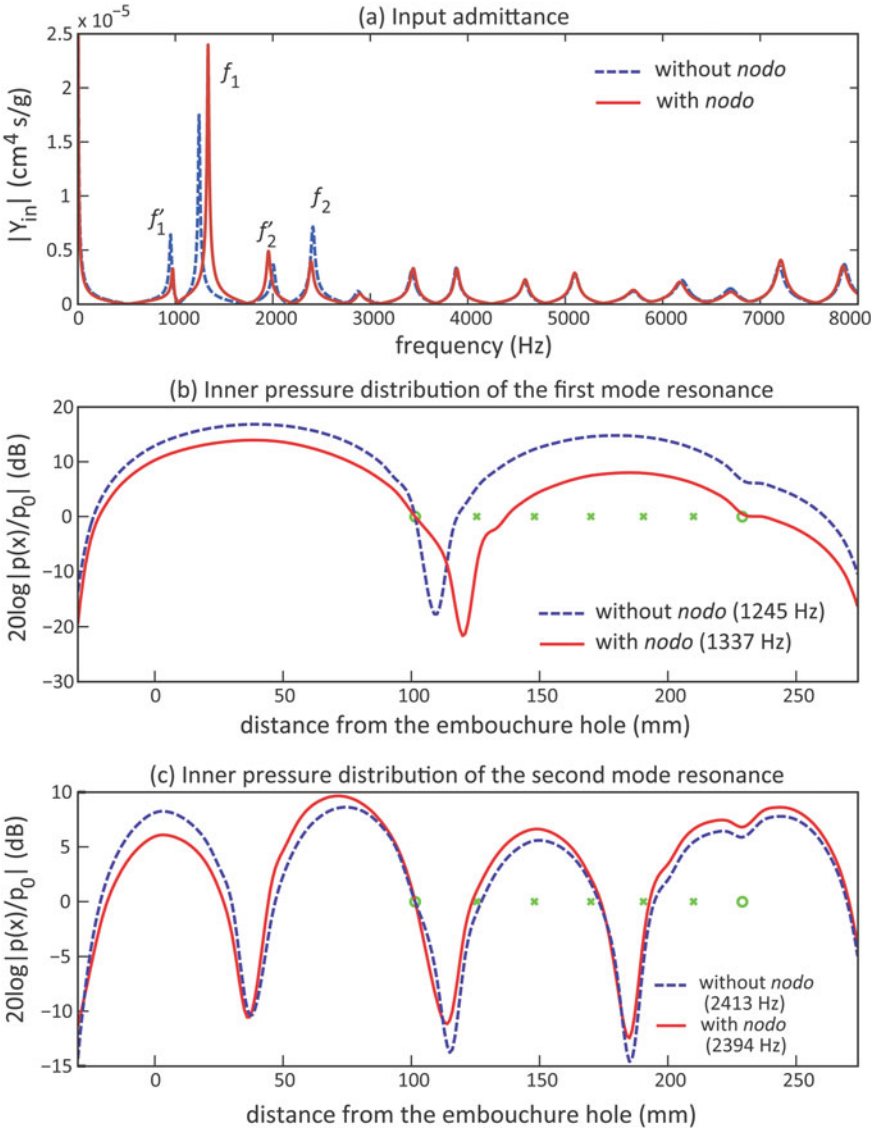


Fig. 18 Results on fingering C [41]. **a** Input admittance; **b** and **c** inner pressure distributions along the bore in the first and second mode resonances, respectively

phase shift (the amplitude kink) is observed at the open tone hole just as in the case of the shakuhachi (cf. Figs. 9 and 11).

A peculiar cross fingering C gives a complicated input admittance as depicted in Fig. 18a. The peaks of f_1 and f_2 should be the resonances of the upper bore above the first open tone hole, but the peaks of f'_1 and f'_2 should be the resonances of the

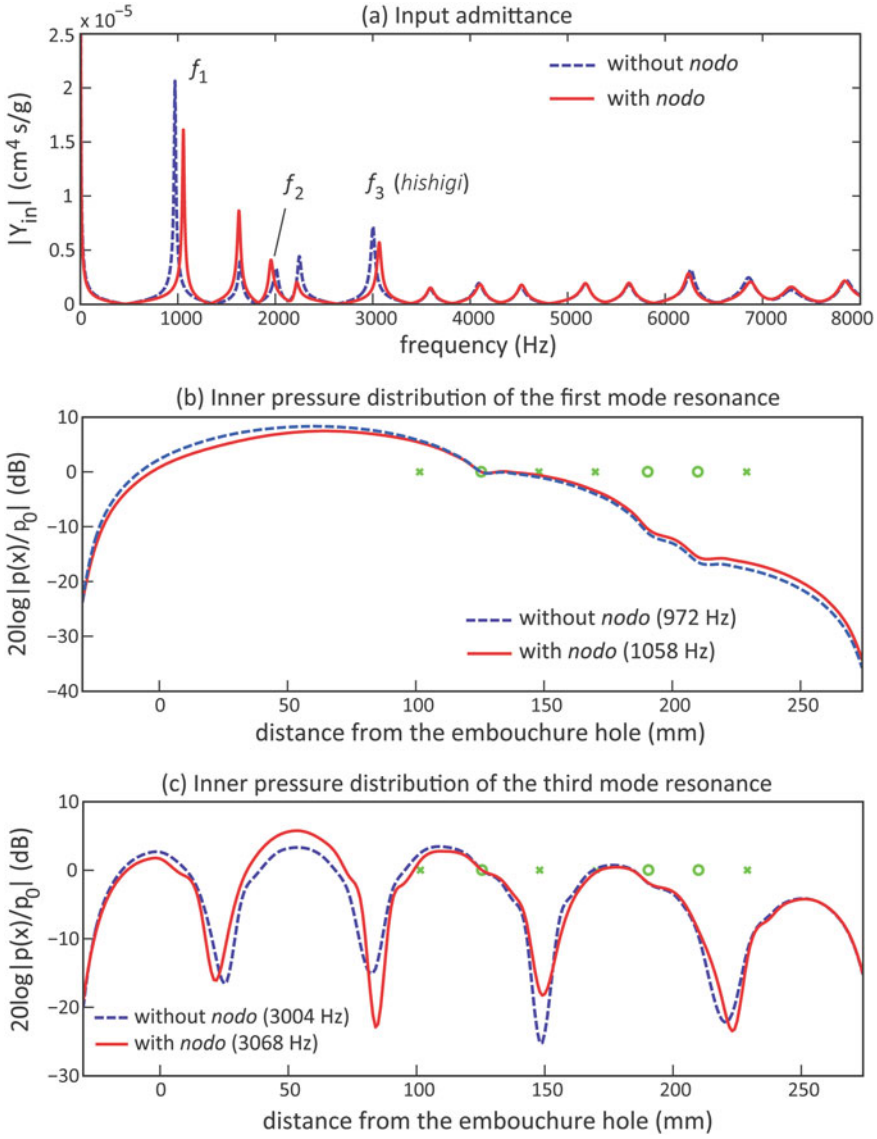


Fig. 19 Results on fingering D [41]. **a** Input admittance; **b** and **c** inner pressure distributions along the bore in the first and second mode resonances, respectively

lower bore below the first open tone hole. The frequency of f_1 is largely ascent by the throat (by 92 Hz). Moreover, f_2 is descent by 19 Hz, and then the octave is definitely shrunk. The inner pressure distribution of the first mode shown in Fig. 18b indicates that similar to Fig. 16c on fingering A. The lower bore is responsible for the pitch sharpening. The dominance of the lower bore is also seen

in Fig. 18c where the longer half-wavelength along the lower bore is observed for the pitch flattening by the throat (from 2413 to 2394 Hz).

As in the case of fingering C, fingering D gives a strong octave shrink by inserting the throat as the result of the ascent in f_1 and the descent in f_2 as shown in Fig. 19a. The admittance peak between f_1 and f_2 might be the first resonance of the lower bore below the second open tone hole, but that between f_2 and f_3 might not be the second resonance of the lower bore. The complicated fingering might cause the bore resonance between the open second and fifth tone holes.

The third mode f_3 of cross fingering D produces the *hishigi* tone. There was a conventional explanation that the throat was inserted to make the hishigi tone be played easily. However, the peak of f_3 is a little lowered by the throat as indicated in Fig. 19a. Therefore, the above explanation seems to be inadequate. Nevertheless, we may recognize a larger half-wavelength pressure around x = 50 mm when the throat is inserted (also, this half wavelength is made longer in spite of the pitch ascent) as depicted in Fig. 19c. The third-mode distribution pattern is made more complicated by the throat.

3.5 *Numerical Calculation on the Effects of* Nodo *Shape*

Numerical calculation of the resonance frequency was moreover carried out concerning four patterns of the throat shape depicted in Fig. 15 and its result was compared with the measurement results [20, 38]. The four patterns were referred to as “no throat” (the blue solid line), “normal throat” (the blue dashed line), “widened throat” (the green solid line), and “narrowed throat” (the red solid line). Eight common fingerings depicted in Fig. 20 were used for our calculation [41, 42], where the end correction at the embouchure hole was fixed to 30 mm for the simplicity. These fingerings were selected from 15 fingerings given in Fig. 5 of Ref. [39]. Each fingering is shown below the staff notation of the resulting tones, which

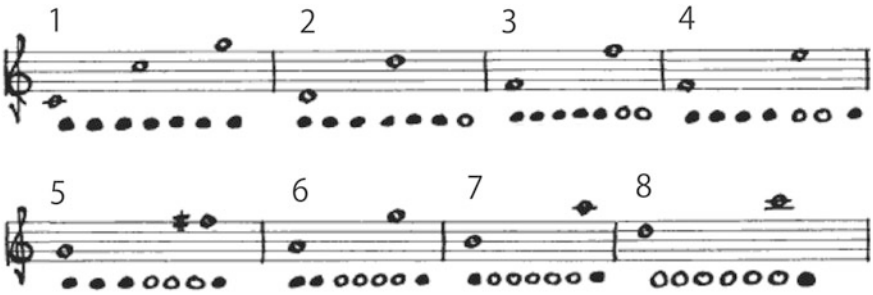


Fig. 20 Eight common fingerings of the nohkan and the resulting approximate tones. The tone hole is opened from the bore bottom in sequence and the tonal pitch is correspondingly ascent. This figure is based on Ref. [39]

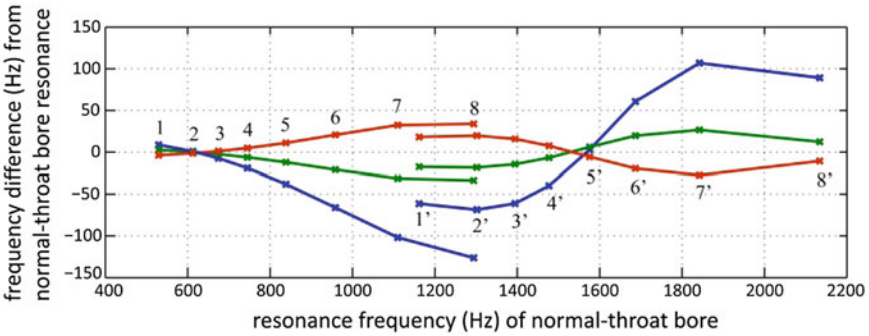


Fig. 21 The resonance frequency differences from the normal-throat resonance frequency [41]. The *red line* a narrowed throat; the *green line* a widened throat; the *blue line* no throat. The *numeral on the curve* denotes the first-mode resonance of each fingering; the *numeral with the apostrophe on the curve* the second-mode resonance of each fingering

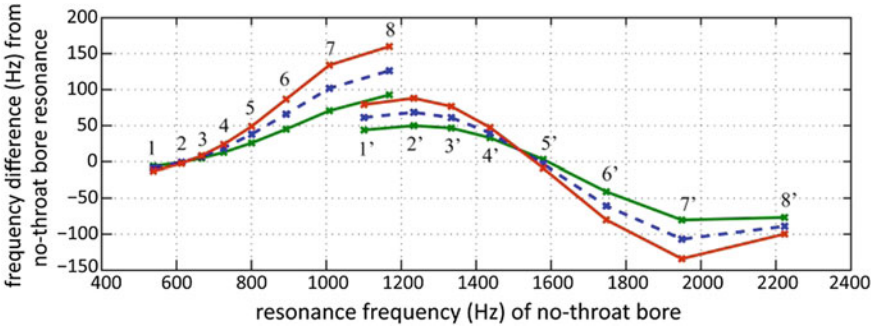


Fig. 22 The resonance frequency differences from the no-throat resonance frequency [41]. The *red line* a narrowed throat; the *green line* a widened throat; the *blue dashed line* normal throat. The *numeral* indicates the fingering; *the apostrophe* the second-mode resonance

only indicate approximate tones. The filled circle denotes a closed tone hole while the open circle denotes an open tone hole.

The calculated result is illustrated in two ways [41]: The abscissa in Fig. 21 is the resonance frequencies in the normal throat case; the abscissa in Fig. 22 is the resonance frequencies in the no-throat case. As shown in Fig. 21, the resonance frequencies of the narrowed-throat and widened-throat bores indicate symmetrical frequency differences from the resonance frequency of the normal-throat bore. In other words, the positive and negative of their frequency differences are reverse each other. The cross-over occurs at fingering 2 in the first register, and it occurs at fingering 5 in the second register. The degree of frequency difference is strengthened as the fingering changes from 2 to 8 in the first register. The measurement [38] gave the cross-over at fingering 4 (near 700 Hz) in the first register, and the cross-over at fingering 7 (near 1900 Hz) in the second register. These disagreements are probably

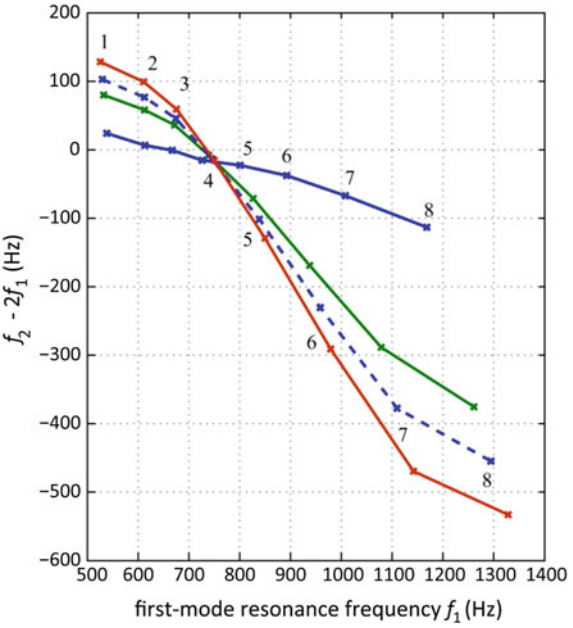


Fig. 23 The octave relation between the first-mode and second-mode resonances [41]. The f_1 denotes the first-mode resonance frequency; f_2 the second-mode one. The *blue solid line* no throat; the *blue dashed line* normal throat; the *green solid line* widened throat; the *red solid line* narrowed throat. The *numeral on the curve* denotes the fingering in Fig. 20

due to the difference between the actual nohkan used for the measurement and the modeled nohkan used for the calculation. The no-throat bore strongly exaggerates the resonance frequencies of the widened-throat bore.

The difference in resonance frequency for each throat model from that for the no-throat bore is depicted in Fig. 22. Except for fingerings 1 and 2, the resonance frequency is raised more and more as the upper tone hole is opened in the first register. The cross-over occurs at fingering 5 in the second register. It should be noted that the resonance frequency is appreciably raised by the throat from about 600–1600 Hz (between the cross-over in the first register and that in the second register) in comparison with the no-throat bore.

The deviation from the octave relation between the first-mode resonance frequency f_1 and the second-mode resonance frequency f_2 is shown in Fig. 23. The abscissa is f_1; the ordinate is $f_2 - 2f_1$. The numeral on the curve denotes the fingering given in Fig. 20. Fingering 1 indicates positive $f_2 - 2f_1$ for all throat shapes, and the positive degree becomes weak as more tone holes are opened in fingerings 2 and 3. Fingering 4 then brings negative $f_2 - 2f_1$ and the negative degree is strengthened as more tone holes are opened in fingerings 5, 6, 7, and 8. The octave relation $f_2 - 2f_1$ mentioned above occurs in the no-throat nohkan due to the reverse conical bore. However, the throat causes much stronger effect to this octave relation. It is well known that the octave shrink is brought as more tone holes are opened in fingerings 4, 5, 6, 7, and 8. This tendency is recognized in Fig. 14 on many cross fingerings, too. Also, our calculation result generally agrees with the measurement result by Ando [20, 38], but the cross-over occurred near 800 Hz or fingering 5.

3.6 Perturbation Theory Applied to the Nohkan

Since the throat can be recognized as a perturbation to the bore, the resonance frequency shift due to the throat may be calculated from the perturbation theory developed by Rayleigh [9, 28]. The following definitions are made: The bore length from the embouchure hole is denoted as x; the pressure distribution along the bore with no throat as $p_0(x)$; the sound speed in the bore as c; the widening of the bore at $x = x_0$ as Δ; the resonance angular frequency of the bore with no throat as ω_0; the bore cross section with no perturbation as $S_0(x)$; the integral of $S_0(x)p_0^2(x)$ along the bore as N. Then, the deviation of the resonance angular frequency $\delta\omega$ due to the positive bore perturbation $\Delta(>0)$ concentrated at $x = x_0$ is given as

$$\delta\omega = \frac{c^2\Delta}{2\omega_0 N}\left[\frac{d}{dx}\left(p_0\frac{dp_0}{dx}\right)\right] \quad at \quad x = x_0 \tag{1}$$

according to Eq. (8.72) in Ref. [9].

If the pressure with no perturbation $p_0 = \sin(2\pi x/\lambda)$ (λ denotes the wavelength and x involves the end correction at the embouchure hole) is inserted in Eq. (1), the quantity in the angular parenthesis is $(2\pi/\lambda)^2 \cos(4\pi x_0/\lambda)$. The value of x_0 may be represented by the middle position (50 mm in Fig. 15) at which the throat is most narrowed and by the assumed end correction at the embouchure hole (30 mm). Hence, $x_0 = 80$ mm is supposed. Since λ, c, ω_0, and N are positive, but Δ is negative for the nohkan throat, the value of the index

$$W = -\cos(4\pi x_0/\lambda) \tag{2}$$

determines the sign of the perturbed frequency $\delta\omega$ of Eq. (1). If λ is converted by c/f ($c = 344$ m/s), Eq. (2) is plotted as Fig. 24 against the resonance frequency f of the nohkan without the throat.

The frequency f_{0n} satisfying $W = 0$ is given by

$$f_{0n} = (2n-1)c/8x_0 (n = 1, 2, 3, \ldots), \tag{3}$$

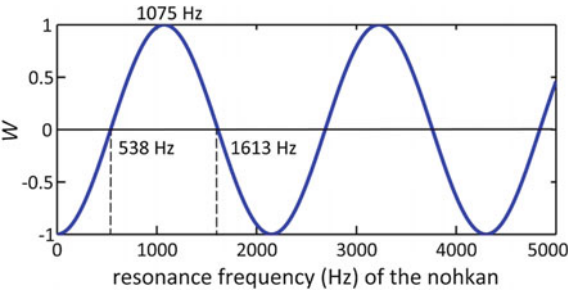


Fig. 24 The deviation of the nohkan resonance frequency expressed by the index W [42]. The bore perturbation Δ given by the normal throat is represented at a point $x = x_0 = 80$ mm including the assumed embouchure end correction 30 mm (cf. Fig. 15)

where n denotes the mode number of the resonance. The values of f_{01} and f_{02} are calculated as 538 and 1613 Hz, respectively. These values match well those given from Fig. 22 which illustrates the resonance frequency difference of the nohkan with the throat from the no-throat nohkan resonance frequency. The dependence of the throat effect on the resonance frequency can be estimated by the perturbation method. Also, as the middle position of the throat is closer to the embouchure hole, the f_{0n} value for each mode n is increased as known from Eq. (3). In other words, the intonation and octave relation in the nohkan can be adjusted not only by the throat shape but also by the throat position [41, 42].

3.7 A Comparison of the Nohkan with the Piccolo

Both the piccolo and the nohkan have a reverse conical bore as shown in Fig. 25 and have a very similar playing range [D_5 (587 Hz) to C_8 (4186 Hz) and C_5 (523 Hz) to F_7 (2794 Hz), respectively]. However, the throat breaks down the octave balance in the nohkan as discussed so far, while the cylindrical headjoint serves to hold the octave balance in the piccolo. The frequency characteristic of the reverse conical bore (without the throat) was already shown in Fig. 16a on fingering A (all tone holes are closed). In this subsection, a cross fingering with two extremely distant open tone holes just like cross fingering C of the nohkan will be discussed between the nohkan and the piccolo. Numerical data on the two tone holes are indicated in Table 1 [41].

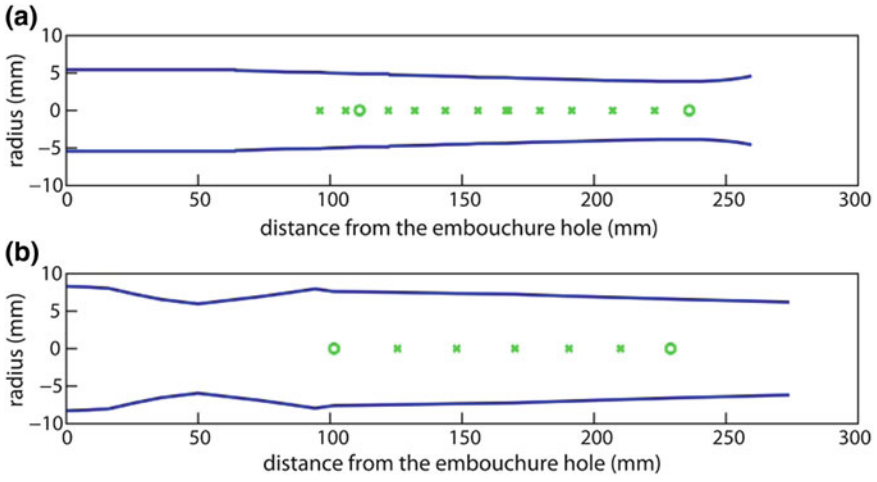


Fig. 25 A cross fingering with two extremely distant open tone holes in a modern piccolo (**a**) and a nohkan (**b**) [41]. The bore radius of (**a**) is based on Ref. [14]

Table 1 The position and diameter of the upper and lower tone holes in a modern piccolo [14] and a nohkan [41]

	Upper tone hole position (mm)	Upper tone hole diameter (mm)	Lower tone hole position (mm)	Lower tone hole diameter (mm)
Piccolo	111.2	4.2	236.1	8.0
Nohkan	101.5	9.2	229.0	8.2

Prior to our comparison between a piccolo and a nohkan, the input admittance $|Y_{in}|$ and the inner pressure distribution $p(x)$ along the bore, which is normalized p_0 at the upper open tone hole, are calculated on Fig. 25a and depicted in Fig. 26. Since this fingering usually gives $D^{\#}{}_6$ (1245 Hz), the second peak in Fig. 26a corresponds to this tonal pitch and the pressure minimum is observed at the upper open tone hole in Fig. 26b on the first mode resonance. Also, it should be noted that a similar but a little different half-wavelength distribution is observed between the two tone holes. Probably the first peak in Fig. 26a is on the first mode resonance f'_1 of the bore below the upper open tone hole as indicated in Fig. 18a. As a result, the piccolo as well as the nohkan can bring about a pair of f_n and f'_n (n = 1, 2, 3, …).

For the simplicity, a common cylindrical pipe instead of actual bores depicted in Fig. 25 is used for our calculation to estimate the acoustical role of the two distinct tone holes. This pipe is 265 mm long and 12 mm in diameter. The two tone holes are located at 100 and 230 mm distant from the embouchure hole respectively, and

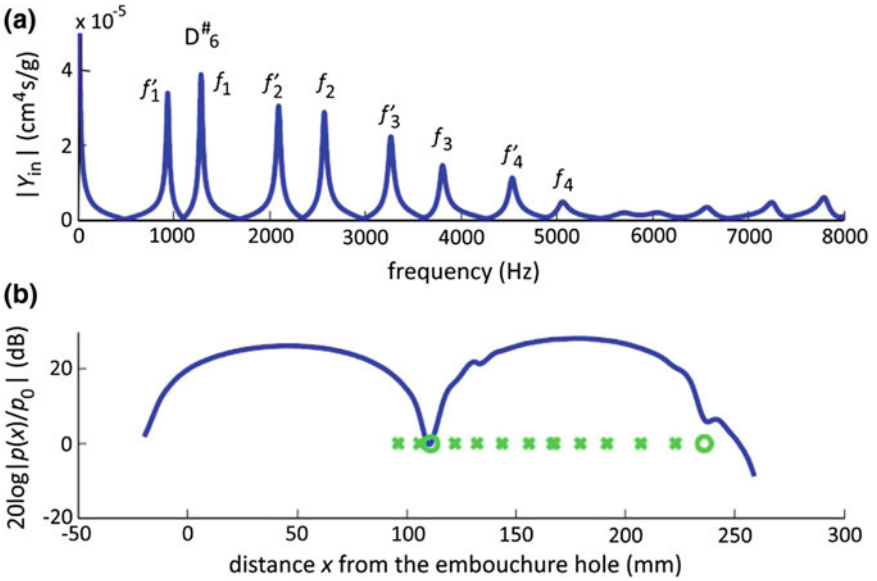


Fig. 26 The calculated input admittance (**a**) and internal pressure distribution along the bore (**b**) of the piccolo on the fingering for $D^{\#}{}_6$ (cf. Fig. 25a) [41]

they are commonly 8 mm in diameter and 3 mm long. The end correction at the embouchure hole is assumed to be 30 mm. On this reference pipe model the input admittance $|Y_{in}|$ and the inner pressure distribution $p(x)$ along the pipe are shown in Fig. 27, respectively [41]. Although six (or three pairing) peaks are shown in frame (a), the pressure distributions of the third modes (f) and (g) are not shown. The lower first mode frequency (1021 Hz) gives the acoustical pipe length 168 mm (for c = 344 m/s), which is very close to the physical length 165 mm of the pipe below the upper open tone hole. Since this mode seems to radiate no sound, this agreement confirms that the lower first mode corresponds to f'_1 (the first mode of the lower bore below the upper open tone hole) in Fig. 26a. On the other hand, the higher first mode (1247 Hz) gives the acoustical pipe length 138 mm, which probably consists of the physical length 100 mm of the upper pipe, the end correction 30 mm

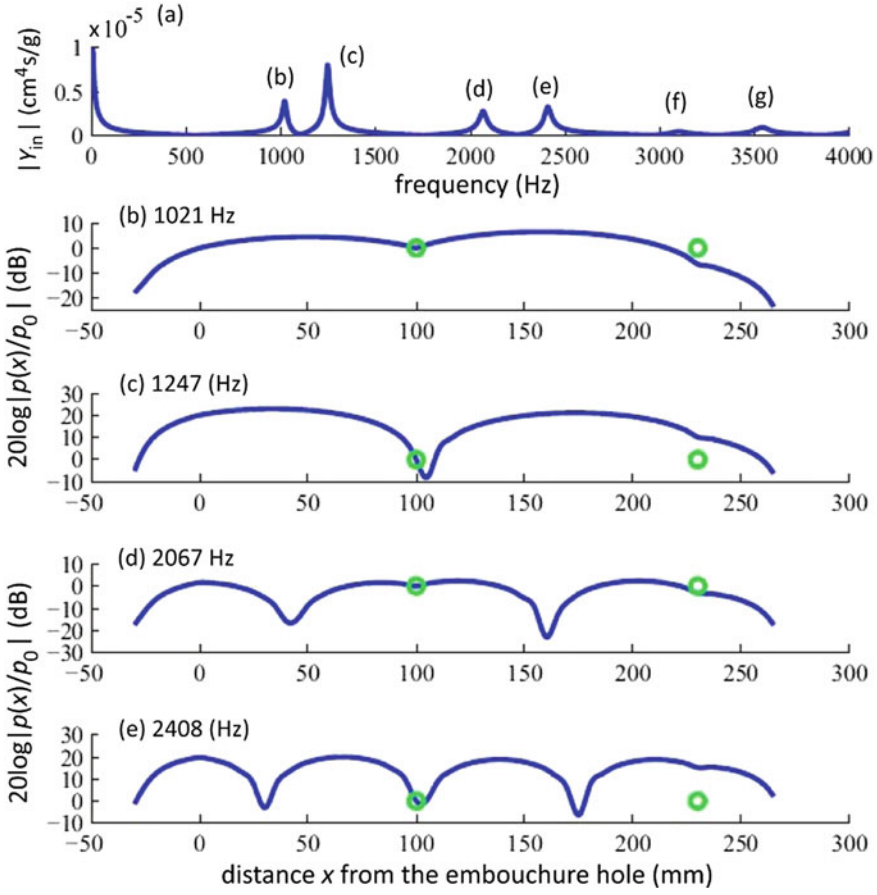


Fig. 27 Input admittance (**a**) and internal pressure distributions of the admittance peaks denoted as (**b**–**e**) in frame **a** concerning the reference cylindrical pipe of d = 8 mm [41]

assumed beforehand at the embouchure hole, and the end correction 8 mm anticipated at the upper open tone hole. This result confirms that the higher first mode corresponds to f_1 (the first mode of the upper bore) in Fig. 26a. The pressure distributions depicted in Figs. 27c, e well indicate the patterns of the first and second resonance modes of the upper pipe above the upper open tone hole, respectively.

Also, another interpretation seems to be possible: The upper open tone hole may operate as an octave hole for the fingering that only the lower hole is opened. Such an interpretation can be derived from much smaller diameter (4.2 mm) of the upper tone hole in the piccolo as shown in Table 1. Therefore, it seems to be worthy to calculate the models with smaller diameters of the upper tone hole. The diameter d is changed from 8 mm to 4, 2, and 0 mm. The corresponding results are shown in Figs. 28, 29 and 30, respectively [41].

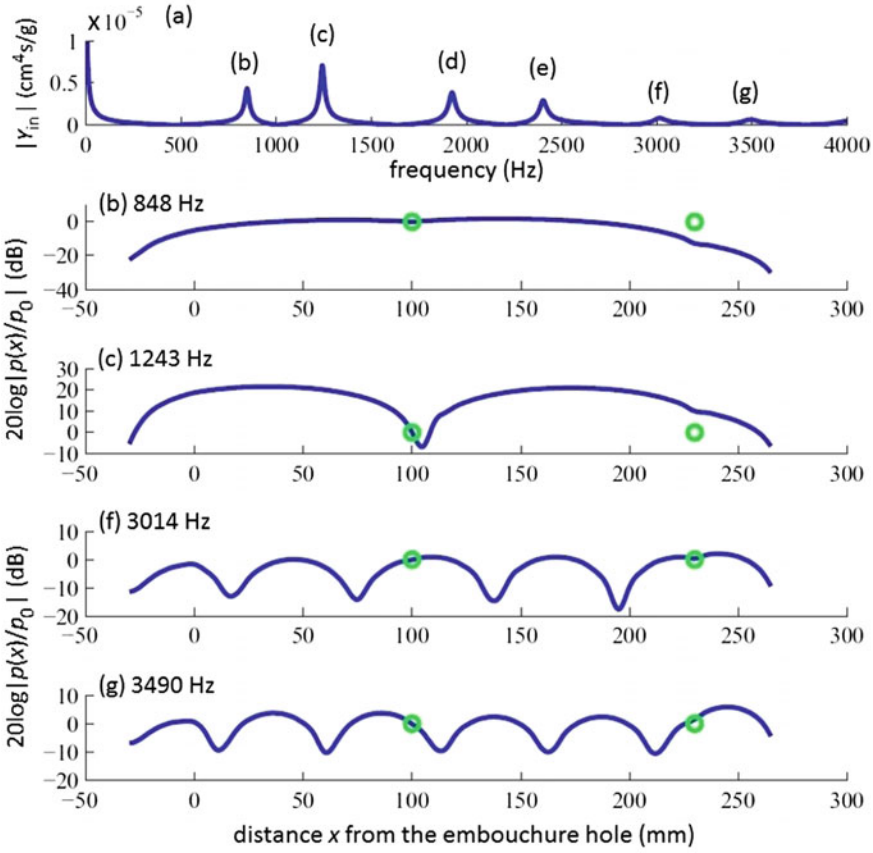


Fig. 28 Input admittance (**a**) and internal pressure distributions of the admittance peaks denoted as (**b**, **c**, **f**, and **g**) in frame **a** concerning the cylindrical pipe with an upper open tone hole of diameter d = 4 mm [41]

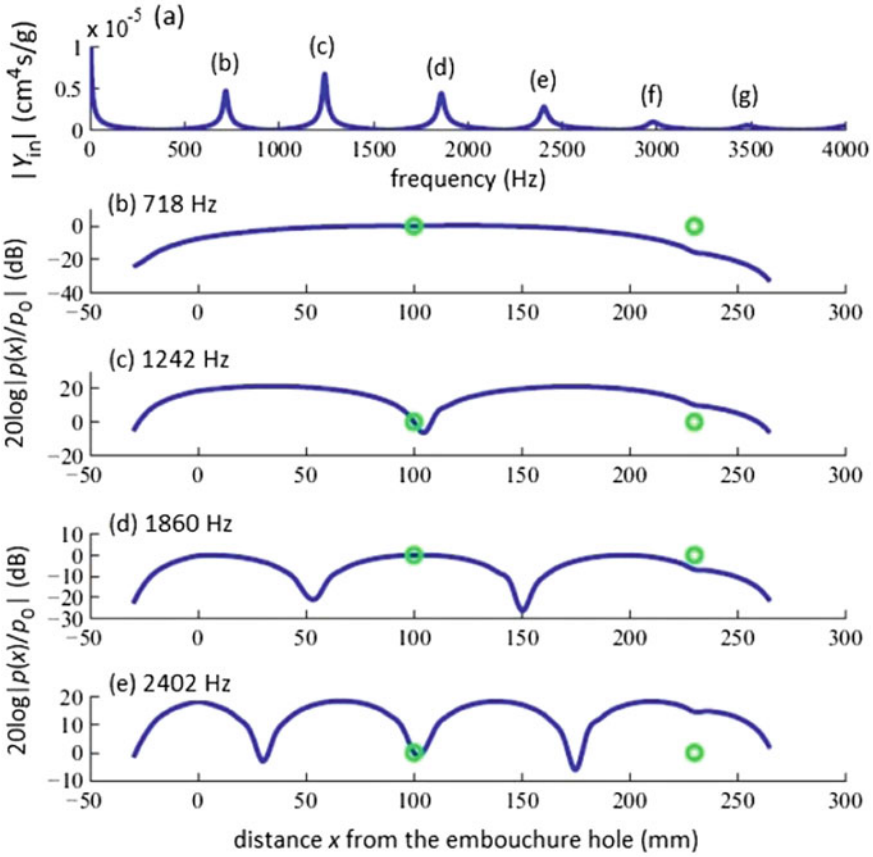


Fig. 29 Input admittance (**a**) and internal pressure distributions of the admittance peaks denoted as (**b**–**e**) in frame **a** concerning the cylindrical pipe with an upper open tone hole of diameter d = 2 mm [41]

It should be noted that the frequencies of admittance peaks (c) and (e) are almost unchanged in Figs. 27, 28, 29 and 30 [in Fig. 28 pressure distributions of (f) and (g) are depicted instead of (d) and (e)]: The frequency of peak (c) varies as 1247, 1243, 1242, and 1243 Hz corresponding to the sequential change in the diameter of the upper tone hole (8, 4, 2, and 0 mm); similarly the frequency of peak (e) varies as 2408, 2403, 2402, and 2405 Hz. If these results are considered based on Fig. 30 concerning the pipe without the upper tone hole, the frequency of peak (b) gradually increases and approaches peak (c) as the upper tone hole becomes larger. Also, at that time the amplitude of peak (b) gradually decreases. According to Fig. 30, peak (b) is understood as the first mode of the pipe when only the lowest tone hole is open. Since the upper tone hole at x = 100 mm is nearly located at the pressure loop of the above first mode, a kink appreciably appears there as the upper tone hole

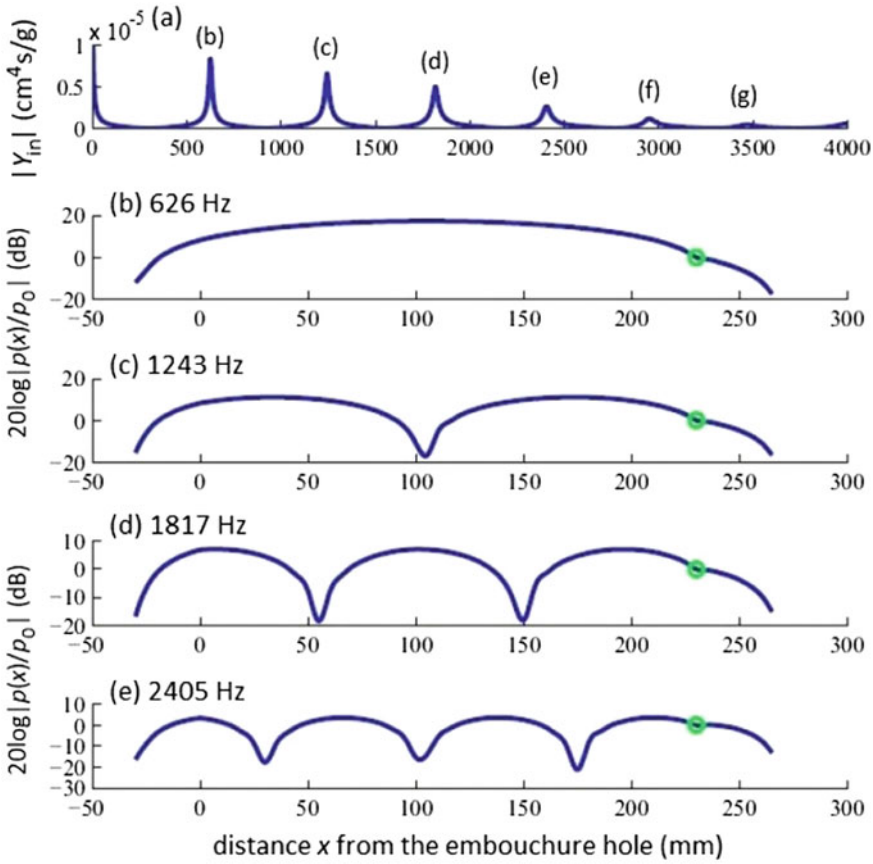


Fig. 30 Input admittance (**a**) and internal pressure distributions of the admittance peaks denoted as (**b**-**e**) in frame **a** concerning the cylindrical pipe with an upper open tone hole of diameter d = 0 mm [41]

is opened larger. Such a tendency is indicated by the pressure distributions on peaks (d) and (f) as shown in Figs. 27, 28 and 29.

Therefore, fingering $D^{\#}_6$ in the piccolo and fingering C in the nohkan, which give pairing admittance peaks, make the second mode of the bore with only the lowest open tone hole produce easier. However, in exchange for this octave-hole effect, the original odd modes [(b), (d), and (f) in the input admittance plot of Fig. 30] are made almost difficult to play because these modes are divided by the upper open tone hole [41]. Moreover, it is known that the original odd modes are completely disappeared when the diameter of the upper tone hole becomes 18 to 19 mm [41]. This calculated result suggests that an open tone hole with a diameter more than 1.5 times a pipe diameter (12 mm) can yield a pipe cut down at the tone-hole position.

4 The Shinobue

Another transverse bamboo flute, *shinobue*, is always used in *matsuri* (Japanese festival for celebrating the gods). Also, *matsuri* is the prayer of common people for their happiness, health, safe, prosperity, and so on. The *sinobue* sounds represent such prayer of common people.

4.1 Brief History

The Japanese *yokobue* is roughly categorized as *ryuteki* for the court ensemble, *nohkan* for the noh play, and *shinobue* for the events performed by common people. Although the shinobue has been exclusively used as a melody instrument for *matsuri* in ensemble with drums, it has been used for other cultural events such as *nagauta* and *kabuki* in ensemble with *shamisen* and singing after it was improved in tuning. Historically, the shinobue has six or seven tone holes, however, the seven tone-hole shinobue is generally used with singing. Since the shinobue is the flute for common people, a great variety of the shinobue with different length and structure are found throughout Japan. It is, therefore, very important to recognize that the shinobue strongly reflects the local color. Even its pitch and intonation are different between local customs [1].

The shinobue has been considered as a simplified version of the ryuteki after it became popular among common people around ninth century. However, there are two kinds of shinobue with six or seven tone holes. Also, the musical scale and the inner structure of the ryuteki and shinobue are different from each other. The unified view on the origin and transfiguration of Japanese *fue* has not been established yet as mentioned about the nohkan in Sect. 3.1.

4.2 Unique Structural Properties

The external view of a shinobue with six tone holes (pitched in around C_5 = 523 Hz) is shown in Fig. 31. Since the bore is stopped at the left edge of the embouchure hole, the physical length for the bore resonance of the lowest tone is about 313 mm. A small hole near the bore bottom is a kind of the ornament hole and it has almost no influence upon the bore resonance in normal playing range except for possible subtle adjustment of the intonation. The shinobue for the festival usually has tone holes with an equal size and with an equal hole-to-hole distance for the easiness in play and production. The hole-to-hole distance in the one shown in Fig. 31 is about 23 mm, but hole sizes are slightly different from each other (the diameter of the third and fourth holes is larger: 8.2 mm long and 7.2 mm wide).

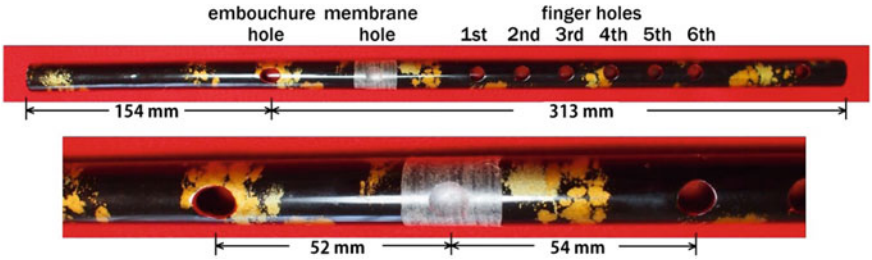


Fig. 31 External view of a shinobue (offered by Shiori Ide). The *upper* and *lower frames* show the total view and the close-up around the membrane hole halfway between the embouchure hole and the first tone hole, respectively

The embouchure hole and tone holes are simply opened and flat against the external surface of the instrument. This is significantly different from the nohkan shown in Fig. 12. Also, both holes are quite smaller and the tone-hole shape is rather close to the circle. These characteristics reflect the blowing and fingering ways different from those in the nohkan. Particularly, *uchiyubi* (the quick finger striking) on a closed tone hole is frequently used in the shinobue.

The bore of the shinobue shown in Fig. 31 is not reverse-cylindrical as in the ryuteki and nohkan. The diameter is 13.4 mm at the embouchure hole; about 12.3 mm at the membrane hole, the first, and second tone holes; 13.2 mm at the third and fourth tone holes; 12.9 mm at the fifth tone hole; about 12.6 mm at the sixth tone hole, the ornament hole and the bore bottom. A wider bore around the larger third and fourth tone holes is characteristic, and it might suggest the importance of these tone holes.

As shown in the lower figure of Fig. 31, this shinobue has a unique hole over which a piece of very thin membrane (traditionally prepared from the inner skin of the bamboo node, called *chikushi*, meaning "bamboo paper") is glued. Although this kind of shinobue with a *membrane hole* is now very rare in Japan, it has been used in northern Kyushu districts. The shinobue in Fig. 31 is still used for the fork festivals called *Kunchi* (meaning *Ku no Hi,* special 9th day in September [43]) in Karatsu, Saga prefecture. This kind of folklore flute with a membrane hole in north Kyushu may be a living evidence of Chinese music influence [44].

The vibrating membrane glued over a hole just halfway between the embouchure hole and the first tone hole (as indicated in Fig. 31) has been used for the *di* (or *dizi*) in China and for the *taegum* in Korea [35, 44]. Transverse flutes with a membrane hole are also played in Mongolia [35]. Therefore, this vibrating membrane probably reflects the tonal taste in East Asia. In China most *di* players prefer a membrane taken from the inner side of a reed because of its even thickness, soft and high elastic quality that creates desired resonant sounds [44]. However, it should be noticed that there is a definite difference between the tonal tastes in the *di* and the *shinobue*: The *di* aims strong and bright reverberant sounds, while the *shinobue* aims beautiful and clear distinguished sounds. See Ref. [44] for the origin of the

membrane hole and historical backgrounds of the *di*, *taegum*, and shinobue with a membrane hole.

The decisive effect of the membrane hole seems to be the high-frequency emphasis and the generation of inharmonic or noise frequencies due to the vibration of the membrane. This tonal effect is common to the effect due to the *sawari* (gentle touch with a string) device which was first applied to the Japanese stringed instruments, *biwa* and *shamisen* [20, 27, 45–47]. The tonal effects brought by the *sawari* and by the membrane hole (probably developed in China and Korea independently) seem to have strong interrelation with the performing environment such as an open-air theater, village square, and *matsuri* procession. In this section the distinctive effects of the membrane hole will be briefly discussed.

4.3 Sound Examples

The tonal difference of the shinobue with and without the membrane hole is shown by the waveforms in Fig. 32a, b, respectively. Tone (a) was played by closing the first and second tone holes in the first register; tone (b) was played by closing the membrane hole in addition to the first and second tone holes. The fundamental frequencies of (a) and (b) were 912.5 and 925.9 Hz, respectively. These tones were a little lower than B^{b}_{5} (932 Hz).

Since the membrane hole approximately locates at the middle of the bore resonating between the embouchure hole (with the assumed end correction of about 20 mm) and the second tone hole, the pressure maximum of the first mode can be formed near the membrane hole, which brings positive perturbation (local

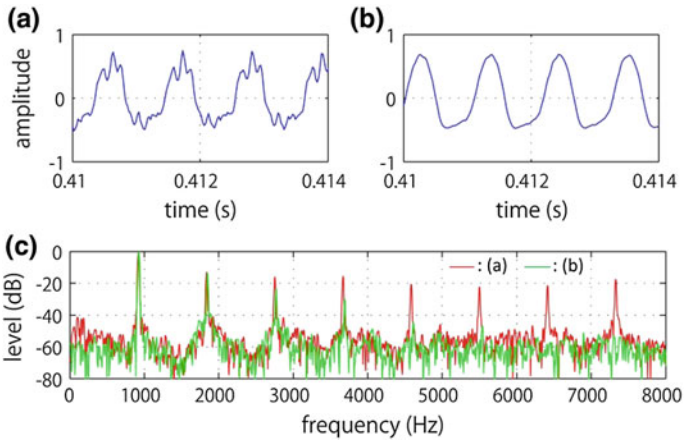


Fig. 32 Sound examples of the shinobue (closely pitched in C_5) played by closing the first and second tone holes in the first register. **a** The waveform when the membrane hole is covered by *chikushi*, the inner skin of the bamboo node; **b** the waveform when the membrane hole is closed; **c** frequency spectra of tones **a** and **b**

cross-sectional enlargement) to the bore. As a result, small frequency lowering occurs in the above tone of the shinobue with the membrane hole (cf. Sect. 3.6).

The waveform of (a) indicates that complicated corrugations are overlapped on a simple waveform of (b). Their frequency spectra are given in Fig. 32c. The harmonics of tone (b) without the membrane hole are poor and almost limited to the fourth harmonic. While all the harmonics of tone (a) maintain almost the same high level up to the eighth harmonic. This high-frequency emphasis seems to be derived from the membrane hole.

4.4 Acoustical Effects of a Membrane Hole

The effect of the membrane hole on the resonance frequency was already explained on the basis of the perturbation theory in the previous subsection. However, the tension, density, and quality of the membrane material as well as the size and location of the membrane hole will affect the instrument sounds. Therefore, an analytic model incorporating these parameters is desired. Samejima [48] and Ide [49, 50] proposed a vibro-acoustical model of the flute with a membrane hole. Their numerical method is based on the coupling between the membrane vibration and the surrounding sound field. Since its description is too complicated to the readers of this book, a brief explanation of the membrane tension effect on a modeled tube is given below.

A tube model is 250 mm long and 40 mm in diameter. Both ends are open. An open hole (18 mm in diameter) is perforated at the position of 140 mm from the top (left) open end. A piece of bamboo paper, whose aerial density is assumed to be 0.00374 (kg/m^2) from the measurement, is glued over this hole. This model tube was used in acoustical measurements [49, 50]. However, it was needed to setup the flat membrane surface for the calculation based on the finite element method, and then the model tube was approximated by a hexagonal column. Numerical calculation on this column with a membrane hole was carried out by changing the membrane tension from 15 to 50 (N/m) stepwise. It should be noted that this applied tension is very low compared with the timpani head to which the tension of 4000–5000 (N/m) is usually applied.

The calculated result is illustrated in Fig. 33, where the dashed line gives the result when the membrane is rigid or when the membrane hole is completely closed [48, 49]. The ordinate, relative SPL (sound pressure level), corresponds to the frequency response function (giving the resonance characteristics) defined by the ratio of the sound pressure at a receiving point (located at the center of the bottom end) to that at the sound-source point (located at the upper side of the top end). It is normalized by the value at the response peak.

Although the model tube with a rigid membrane hole indicates almost harmonic structure as shown by the dashed line (the peak frequencies are 628, 1268, 1892, 2546, 3180, 3808, and 4496 Hz in sequence), the model tube with a vibrating membrane causes appreciably inharmonic structure as shown by the solid line in

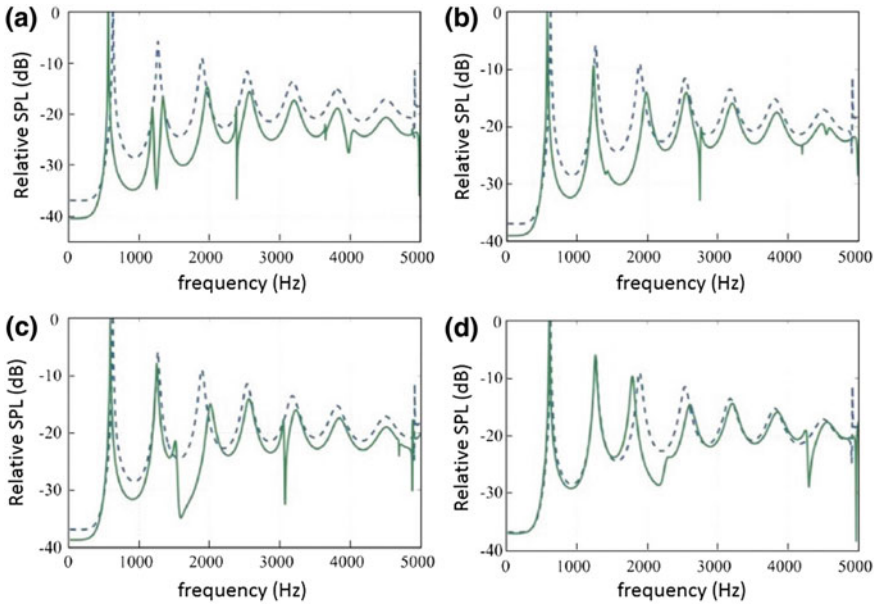


Fig. 33 Calculated effects of the membrane tension T on the frequency response function of a simplified model consisting of a hexagonal column with open ends and a membrane hole [48, 49]. **a** T = 15 N/m; **b** T = 20 N/m; **c** T = 25 N/m; **d** T = 50 N/m. The *dashed line* in each frame indicates the result when the membrane hole is completely closed

each frame of Fig. 33. However, this inharmonicity seems to be diminished as the membrane tension T is increased. Frame (d) for T = 50 N/m shows the harmonic second mode.

The fundamental frequency f_1 in each frame of Fig. 33 approaches to that when the membrane hole is completely closed. The f_1 for T = 15 and 25 N/m is 564 and 592 Hz, respectively; f_1 for $T = \infty$ is 628 Hz. This result qualitatively corresponds to the sound example shown in Fig. 32, where f_1 for the finite T is 912.5 Hz and f_1 for the infinite T is 925.9 Hz. Probably it may be assumed that the tension of 15–25 N/m is a little too weak in comparison with the actual case.

It is interesting that the resonance peaks on the dashed line split into two due to the membrane vibration. It is very appreciable in Fig. 33a, where the second, fourth, and sixth resonance peaks on the dashed line split into two. This peak splitting might be interpreted as follows: If the membrane tension is quite weak, the membrane hole can be like an open hole. As a result, the resonance peaks of the upper tube (140 mm long) above the membrane hole and the lower tube (110 mm long) below the membrane hole might be produced. The second peak frequency on the dashed line is 1268 Hz; the corresponding splitting two peak frequencies in Fig. 33a are 1188 and 1440 Hz, respectively. If the end correction is assumed to be 10 mm, the first-mode resonance frequency of the upper and lower tubes is calculated as 1147 and 1433 Hz, respectively. On the other hand, another interpretation might be allowed:

One of two splitting peaks might be derived from the modal frequency of the circular membrane that is considerably lowered by the surrounding air loading [49]. For example, the small third peak (1516 Hz) on the solid line in Fig. 33c is considered as an original (0, 1) mode (4477 Hz) of the membrane vibration in Ref. [49]. More detailed calculation and experiment are necessary to judge which hypothesis is valid.

Also, the variation of the third peak in Fig. 33 should be paid attention. The frequency of the third peak appreciably higher than that on the dashed line and ascents (from 1968 to 2012 Hz) as the membrane tension increases from 15 to 25 N/m. However, that frequency for $T = 50$ N/m is significantly lower than that on the dashed line as shown in Fig. 33d. The reasonable cause of this inversion is unknown. The interaction between the air-column resonance and the membrane vibration, particularly in actual shinobue with tone holes, is an intricate and intriguing topic in musical acoustics, which is hopefully solved in the near future by incorporating the blowing mechanism.

5 Conclusions

Distinctive characteristics of Japanese (or Asian) flutes (*fue*) and their music are recognized in (1) wind noise as an essential part of the instrument tone, (2) timbre-oriented music rather than octave-oriented or harmonic-oriented music, and (3) high-frequency emphasis suitable for the performing environment (particularly, for the outdoors). If there is no wind noise, the shakuhachi and nohkan cannot create their own sounds. The nohkan has been pursuing very solid and hard timbre and has definitely upset the octave expected when overblown by inserting a kind of obstacle, the throat (called *nodo*) between the embouchure hole and the top tone hole. The classical (*komuso*) shakuhachi traditionally prefers personal deeper introspection to musical performance itself. The nohkan used to be played in outdoors and produce very high and penetrating sounds. In *matsuri* (Japanese folk festival) procession, the shinobue with a membrane hole perforated between the embouchure hole and the top tone hole can create the decisive effect of the high-frequency emphasis due to the membrane vibration.

The classical (*komuso*) shakuhachi has been played in religious environments and appreciated natural bamboo structure. Its original construction had no *ground-paste finish* as seen in the modern shakuhachi but had very thin *urushi* coating over the inner bore surface. The diaphragms were not completely removed and small ridges were retained on the inside bamboo nodes. These remaining diaphragms might affect the intonation and timbre, although their effects are not satisfactorily elucidated yet. The essential difference between classical and modern shakuhachis is in the difference of the inner bore shape with or without the remaining diaphragms. The calculation of the input admittance is very effective to make the difference clear.

The wind noise is dominant between about 1.5–3 kHz when the classical shakuhachi is blown by *muraiki* (a rough and strong blow) during the starting transient.

The definite contrast between noisy, rough sound by *muraiki* and clear, almost noiseless sound by smooth blowing is the essence in shakuhachi sounds.

Also, cross fingerings in the shakuhachi playing, which yield subtle timbre taste, are investigated through the measurement and calculation of the internal standing-wave patterns and the calculation of the input admittance. Cross fingerings in the shakuhachi often causes pitch sharpening (called *intonation anomaly* in this chapter) instead of conventional pitch flattening. This is mainly due to only five tone holes in the shakuhachi. The intonation anomaly is generated if the resonance in the upper bore above the top open tone hole or in the lower bore below the top open tone hole leads the resonance of the whole bore. In other words, the complete coupling between the upper and lower bores through an open top open tone hole is derived when both bores give very close resonance frequencies in the fundamental or higher harmonic modes. The intonation anomaly strongly depends on the position of the top open tone hole. This intonation anomaly has been first quantitatively described through the shakuhachi acoustics.

A special tube device (about 80 mm long), the throat (*nodo*), is intentionally inserted between the embouchure hole and the top tone hole in the nohkan. As a result, the nohkan bore becomes narrowest near the middle of the throat, and then the octave relation between the first and second registers is upset. This throat shape in the nohkan is a great contrast to the cylindrical headjoint in the piccolo that yields the correct octave, although the both have a reverse (convergent) conical bore below the throat and the headjoint, respectively. Acoustical effects of the throat are investigated on the bore with or without the throat and on the bores with different throat shapes through the calculations of the input admittance and the internal pressure distribution for several fingerings. The throat makes the notes of the second octave increasingly flat to the lower octave as the scale is ascent. This octave shrink is generally carried out by the major increase in the first-mode resonance frequency and the minor decrease in the second-mode resonance frequency. The calculated result shows good agreement with the conventional measured result if the fact that the nohkan has no standard length and no consistent pitch is considered. Also, the admittance suggests that the throat does not operate to play easier the *hishigi* tone, the powerful piecing altissimo tone. Moreover, a comparison between the nohkan and the piccolo on a cross fingering with two extremely distant open tone holes reveals that the upper tone hole operates as an octave hole.

The spectra of the first-register tones in the nohkan usually show indefinite harmonic structure. The lower harmonics are seemingly masked by wind noise from about 1 to 2.5 kHz (a little lower than that in the shakuhachi). However, higher harmonics can be usually recognized in the second-register tones of the low-pitched fingerings. On the other hand, the second-register and third-register tones (above 2 kHz) of the high-pitched cross fingerings produce almost sinusoidal waveforms. These tonal differences in the registers and fingerings are very characteristic in the nohkan. Through this tonal tendency and construction method based on the *urushi* coating, the nohkan has been always pursued very solid and hard timbre for the high-pitched tones.

The shinobue has been played by common people in various occasions, particularly in local festivals, *matsuri*. The pitch, intonation, and tone itself of the shinobue largely depend on the locality. The shinobue played in the north Kyushu has been kept the original one traditionally played in China and Korea. It has a membrane hole just halfway between the embouchure hole and the top tone hole. The membrane vibration, which is affected by the hole position, membrane material, membrane tension, and so on, produces characteristic high-frequency emphasis, which should be very effective in *matsuri* procession. This phenomenon is common to the stringed instruments with the *sawari* (gentle touch) mechanism against the string vibration. Therefore, the high-frequency emphasis seems to be the tonal taste common to musical instruments in East Asia.

Bibliography

1. Shibata, M.: Listening to Japanese Sounds. Seidosha, Tokyo, pp. 10–17, 258 (in Japanese) (1987)
2. Shibata, M.: Musical History and Musical Theory. Iwanami, Tokyo, pp. 30–34 (in Japanese) (2014)
3. Malm, W.P.: Traditional Japanese Music and Musical Instruments. Kodansha International, Tokyo Chaps. 4 and 6 (2000) (originally published in 1959 by Charles E. Tuttle Company under the title *Japanese Music and Musical Instruments*)
4. Hayashi, K.: Study of the Shosoin musical instruments. Kazama-Shobo, Tokyo, pp. 164–173 (in Japanese) (1964)
5. Tsukitani, T.: Diffusion and interchange of the Fuke shakuhachi. In Japanese music and Asian music Vol. 3, Propagation and Transfiguration, Iwanami, Tokyo, pp. 37–42 (in Japanese) (1988)
6. Yoshikawa, S.: Changes in acoustical design from ancient shakuhachi to modern shakuhachi. Proceedings of Acoustics 2012 Hong Kong, 2aMU4. Hong Kong (2012)
7. Fletcher, N.H., Douglas, L.M.: Harmonic generation in organ pipes, recorders, and flutes. J. Acoust. Soc. Am. **68**, 767–771 (1980)
8. Yoshikawa, S.: Harmonic generation mechanism in organ pipes. J. Acoust. Soc. Jpn. (E) **5**, 17–29 (1984)
9. Fletcher, N.H., Rossing, T.D.: The Physics of Musical Instruments, 2nd edn. Springer, New York, Chaps. 4, 8, 16 and 17 (1998).
10. Yoshikawa, S.: Jet-wave amplification in organ pipes. J. Acoust. Soc. Am. **103**, 2706–2717 (1998)
11. Yoshikawa, S., Tashiro, H., Sakamoto, Y.: Experimental examination of vortex-sound generation in an organ pipe: a proposal of jet vortex-layer formation model. J. Sound Vib. **331**, 2558–2577 (2012)
12. Terada, T.: Acoustical investigation of the Japanese bamboo pipe, syakuhati. J. Coll. Sci. Tokyo **21**, 1–34 (1907)
13. Benade, A.H.: Fundamentals of musical acoustics. Oxford University Press, New York, Chap. 21 (1976).
14. Nederveen, C.J.: Acoustical aspects of woodwind instruments, 2nd edn. Northern Illinois University Press, DeKalb, USA, pp. 50–53, 132–133 (1998)
15. Wolfe, J., Smith, J.: Cutoff frequencies and cross fingerings in baroque, classical, and modern flutes. J. Acoust. Soc. Am. **114**, 2263–2272 (2003)

16. Yoshikawa, S., Kajiwara, K.: Cross fingerings and associated intonation anomaly in the shakuhachi. Acoust. Sci. Tech. **36**, 314–325 (2015)
17. Ando, Y.: Input admittance of shakuhachis and their resonance characteristics in the playing state. J. Acoust. Soc. Jpn. (E) **7**, 99–111 (1986)
18. Ando, Y., Ohyagi, Y.: "Measuring and calculating methods of shakuhachi input admittance. J. Acoust. Soc. Jpn. (E) **6**, 89–101 (1985)
19. Ando, Y., Tsukitani, T., Maeda, M.: On the shakuhachi structure. In: Sumikura, I., Takano, N., Tohkawa, S., Watanabe, K. (ed.) Music and musicology, Ongaku-No-Tomo-Sha, Tokyo, pp. 39–70 (in Japanese) (1986)
20. Ando, Y.: Acoustics of musical instruments, 2nd edn., Ongaku-No-Tomo-Sha, Tokyo, pp. 79–90, 193–203 (in Japanese) (1996)
21. Simura, S.: An Organology of Kokan Syakuhati (Old Pipe Syakuhati). Shuppan-Gejutsu-Sha, Tokyo (in Japanese) (2002)
22. Waltham, C., Yoshikawa, S.: Construction of wooden musical instruments. In: Bader, R. (ed.) Springer Handbook for Systematic Musicology, pp. 2–16. Springer, New York (to be published)
23. Plitnik, G.R., Strong, W.J.: Numerical method for calculating input impedance of the oboe. J. Acoust. Soc. Am. **65**, 816–825 (1979)
24. Keefe, D.H.: Theory of the single woodwind tone hole. J. Acoust. Soc. Am. **72**, 676–687 (1982)
25. Keefe, D.H.: Experiments on the single woodwind tone hole. J. Acoust. Soc. Am. **72**, 688–699 (1982)
26. Yoshikawa, S., Saneyoshi, J.: Feedback excitation mechanism in organ pipes. J. Acoust. Soc. Jpn. (E) **1**, 175–191 (1980)
27. Ando, Y.: Several peculiar structural devices of Japanese musical instruments and some aspects of timbre taste in Japanese music common with western ones. J. Acoust. Soc. Jpn. (E) **14**, 377–382 (1993)
28. Rayleigh, L.: The theory of sound, vol. 2. Macmillan, London, reprinted 1945, Dover, New York (1894)
29. Caussé, R., Kergomard, J., Lurton, X.: Input impedance of brass musical instruments: Comparison between experimental and numerical models. J. Acoust. Soc. Am. **75**, 241–254 (1984)
30. Mapes-Riordan, D.: Horn modeling with conical and cylindrical transmission-line elements. J. Audio Eng. Soc. **41**, 471–484 (1993)
31. Ebihara, T., Yoshikawa, S.: Nonlinear effects contributing to hand-stopping tones in a horn. J. Acoust. Soc. Am. **133**, 3094–3106 (2013)
32. Benade, A.H.: On the mathematical theory of woodwind finger holes. J. Acoust. Soc. Am. **32**, 1591–1608 (1960)
33. Lefebvre, A., Scavone, G.P.: Characterization of woodwind instrument toneholes with the finite element method. J. Acoust. Soc. Am. **131**, 3153–3163 (2012)
34. Kajiwara, K., Yoshikawa, S.: Discussion on the internal sound pressure distributions when cross fingerings are used in the shakuhachi. Proc. Autumn Meet. Acoust. Soc. Jpn., 903–906 (in Japanese) (2013)
35. Baines, A.: The Oxford companion to musical instruments, Oxford University Press, New York, pp. 21, 58, 90, and 122 (1992)
36. National Museum of Japanese History (ed.), The origin of Japanese musical instruments. Daiichishobo, Tokyo. Chap. III *Fue* (in Japanese) (1995)
37. Ando, Y.: An acoustical study of "Nokan", Ongakugaku (J. Jpn. Musicol. Soc.) **12**, 143–158 (in Japanese) (1966)
38. Ando, Y.: An acoustical study of Japanese transverse flutes. NHK Tech. Res. **22**, 63–78 (in Japanese) (1970)
39. Berger, D.P.: The nohkan: its construction and music. Ethnomusicology **9**, 221–239 (1965)
40. Yoshikawa, S., Kajiwara, K.: Acoustics of the Japanese noh flute, nohkan. Organological Congress 2014, Braga (2014)

41. Kajiwara, K.: Numerical Calculation Method of Sound Pressure Distributions Along the Bore and its Applications to Woodwind Instruments. Master Thesis, Graduate School of Design, Kyushu University (in Japanese) (2014)
42. Kajiwara, K., Yoshikawa, S.: Numerical calculation on acoustical characteristics of the nohkan. Rep. Musical Acoust. Res. Group, MA2013-66, pp. 59–64 (in Japanese) (2013)
43. Yanagida, K.: Japanese Festivals. Kadokawa, Tokyo, originally published in 1932, p. 54 (in Japanese) (1969)
44. Odaka, A.: The resonance hole with membrane; A distinctive feature of East Asian transverse flutes. Proceedings of Acoustics 2012 Hong Kong, 2aMU1. Hong Kong (2012)
45. Yoshikawa, S.: Acoustical classification of woods for string instruments. J. Acoust. Soc. Am. **122**, 568–573 (2007)
46. Yoshikawa, S., Shinoduka, M., Senda, T.: A comparison of string instruments based on wood properties: Biwa vs. cello. Acoust. Sci. Tech. **29**, 41–50 (2008)
47. Yoshikawa, S.: Chapter 11 "Plucked string instruments in Asia. In: Rossing, T.D. (edn.) The Science of String Instruments. Springer, New York (2010)
48. Samejima, T.,Ide, S., Araki, Y.: Vibro-acoustic analysis of wind instruments with membranes on resonance holes. Proceedings of Acoustics 2012 Hong Kong, 2aMU2. Hong Kong (2012)
49. Ide, S.: Acoustic Characteristics of Wind Instruments with Membrane on Resonance Hole. Master Thesis, Graduate School of Design, Kyushu University (in Japanese) (2011)
50. Ide, S., Samejima, T., Araki, Y.: Numerical coupled analysis and measurement of wind instruments with membranes on resonance holes. Proceedings of Autumn Meet. Acoust. Soc. Jpn., pp. 961–962 (in Japanese) (2011)

Author Biography

Shigeru Yoshikawa After graduating from Physics Department of Nagoya University in 1974, Shigeru **Yoshikawa** started acoustical research on organ pipes. He worked for Technical R&D Institute of Defense Ministry from 1980 and investigated underwater organ pipes while personally studying musical acoustics. He was a professor of musical instrument acoustics at Graduate School of Design, Kyushu University from 1998 and was retired in 2015.

Acoustics of the Qin

Chris Waltham, Kimi Coaldrake, Evert Koster and Yang Lan

Abstract The qin (guqin, chi'in) is a seven-string zither with long roots that run deep into Chinese history. The instrument occupies a central place in Chinese musical culture, one that may be compared to that of the violin in Western music; both instruments are deemed to have attained some level of perfection. The qin's history is intertwined with philosophy and folklore and its construction is replete with symbolism. While much has been written on the evolution, organology and playing practice of the qin, little has been published about its acoustics. This chapter explores the qin's background and summarizes the few acoustical studies that have been published to date. Many questions remain, and prospects for future research are discussed.

1 Introduction

The qin (in pinyin, otherwise *chin* or *ch'in* in the Wade-Giles romanization) is an instrument that occupies an ancient and central place in the musical and philosophical culture of China [1]. It is a plucked seven-string zither and is often referred to as the *guqin*, with the prefix "gu" indicating its great antiquity. Like an old Italian violin, a fine qin is considered an objet d'art as well as a musical instrument.

C. Waltham (✉) · E. Koster · Y. Lan
Department of Physics and Astronomy, University of British Columbia,
Vancouver, BC V6T 1Z1, Canada
e-mail: cew@phas.ubc.ca

E. Koster
e-mail: evko@phas.ubc.ca

Y. Lan
e-mail: yanglan@phas.ubc.ca

K. Coaldrake
Elder Conservatorium of Music, University of Adelaide,
Adelaide, SA 5005, Australia
e-mail: kimi.coaldrake@adelaide.edu.au

A. Schneider (ed.), *Studies in Musical Acoustics and Psychoacoustics*,
Current Research in Systematic Musicology 4,
DOI 10.1007/978-3-319-47292-8_2

The four arts that every Imperial Chinese scholar-gentleman was expected to master were *qin, qi, shu* and *hua* (i.e. music, the game of go, calligraphy and painting), of which the qin came first. Failing mastery, scholar-gentlemen needed at least to own a qin, even if it had no strings [2].

The very name of the qin is linked via homonyms to concepts of rectitude and restraint. Thus the Book of Rites, Liji, which records social norms of the Zhou Dynasty (c.1046–256 BCE) states: "Lute[1] means restraining. With this instrument licentiousness and falsehood are restrained, and the human heart is rectified" (p. 41 in Ref. [3]). The names of each part of the instrument are replete with symbolism and lore.

In the course of this chapter we review two recent publications on qin acoustics. One is a published thesis in Chinese by Yang [4, 5] who did extensive work on finite-element modelling of the qin soundbox (without any fluid coupling) and also the mechanical properties of the strings. The second is a study of radiation and significant radiating modes conducted by three of the present authors [6].

We first explore at little of the organology of the qin, and the physical characteristics of the instrument we know today. The woods used in construction, paulownia and catalpa, are little known in West, and are not used in any common Western instrument. What is known of the mechanical properties of these woods will be summarized. Following that there is a section on strings and tuning, including the properties of silk strings, and the more recently popular silk-wrapped steel strings. Next the vibro-acoustics of the qin soundbox is discussed, including the important effect of the table that is generally used to support the instrument. Lastly, we examine the possibilities opened up by new finite-element software that include fluid-structure coupling.

2 History

The earliest known Chinese zithers, known as *se*, are found in tombs from the middle Spring and Autumn Period (c.771–476 BCE). The *se* typically had 25 strings with movable "flying geese" bridges in the manner of the modern *zheng*. What distinguishes qins from the *se* type of zithers is the single bridge at one end of the soundbox. The earliest known example of a qin was found in the Warring States period tomb of Marquis Yi (c. 433 BCE). This qin had 10 strings and differs from a modern instrument in that the soundbox is only 43 cm long while the overall length, with a fingerboard extension giving a total length of 67 cm, the strings having a playing length of 62–63 cm [7]. In this form the instrument bore some resemblance to a lute, which is why some older texts refer to it as a Chinese lute (e.g. Ref. [3]).

[1]Why it is called a "lute" rather than a zither is explained in Sect. 2.

Also found in Marquis Yi's tomb was a long 5-string zither with a single bridge at one end that looks like a thin version of a modern qin (and of a very similar length—120 cm). However, scholars generally do not regard this zither as a qin and debate whether it was a *junzhong*, an instrument used for tuning bells, or a *zhu*, an instrument played by striking, rather than plucking, the strings [8].

Over the course of the next centuries, the qin slowly lengthened and settled with seven strings by the second century BCE. Its current form was attained around 400 CE [7]. Modern instruments (Fig. 1) closely resemble surviving examples from the Tang dynasty [9, 10], 618–907 CE. For historical instruments, establishing the wood used in construction is mired in confusion of nomenclature as well as the usual problems of identifying wood in museum objects (see p. 13 of Ref. [11]).

The age of a qin can be roughly determined by the craquelure in the lacquer [12]. After a century or two, the pattern of cracks is said to look like a flowing stream or a serpent's belly. After that they take on the appearance of a cow's tail. After three or four centuries, they reach the most desired pattern, that of plum blossoms or tortoise shell, that is the mark of a fine old qin.

3 Construction

The construction of the instrument is described in detail in an 1855 document by a qin maker from Fukien province, Chu Feng-chieh [11]. The qin's length is approximately 1200 mm long (365 *fen*, one for each day of the year: 1 *fen* = 1/100 Chinese foot), 200 mm wide at the bridge end, tapering to 150 mm at the nut ("dragon's gums"), and 50 mm deep (Figs. 1 and 2). The soundbox is made in two shaped halves, the front of modern examples usually being carved out of paulownia (*paulownia tomentosa*, 泡桐, *paotong* in pinyin) or China fir (*cunninghamia lanceolata*, 杉木, *shanmu*) (Ref. [13], p.37). "Pao", 泡 means "porous" and "low density" in Chinese, which aptly describes paulownia wood. *P. tomentosa* and *F. simplex* are collectively known as tongmu, i.e. tong (桐) wood. Note that references to tong wood often appear in older texts in the Wade-Giles form as *t'ung*, leading to confusion with the wood of the tung-oil tree (*Vernicia fordii*, 油桐), which is not used for qin making.

The back half of the qin soundbox is a flat piece of Chinese catalpa (*catalpa ovata*, 梓木, *zimu*) or Chinese cypress (*cupressus L.*, 柏木, *baimu*), each piece being at least 10 mm thick, and flat-sawn. Depending on the wood used, the mass can vary from less than 2 kg to more than 3 kg [6]. The two halves are traditionally bonded with lacquer, which may be mixed with powdered gemstones, powdered deer-horn, clay or plaster [11, 12]. Lieberman [12] states that most cements are mixed with copper dust to obtain "golden, stone-like tones", a claim not easy for an acoustician to verify. However, modern makers use glues of various kinds. There are two sound holes in the back; the one near the centre (the "dragon pool", referred to here as H1) is approximately 200 mm × 30 mm in size, and the one near the nut (the "phoenix pond", referred to here as H2) is approximately 100 mm × 30 mm in

Fig. 1 Front and back of a typical qin (Author photographs)

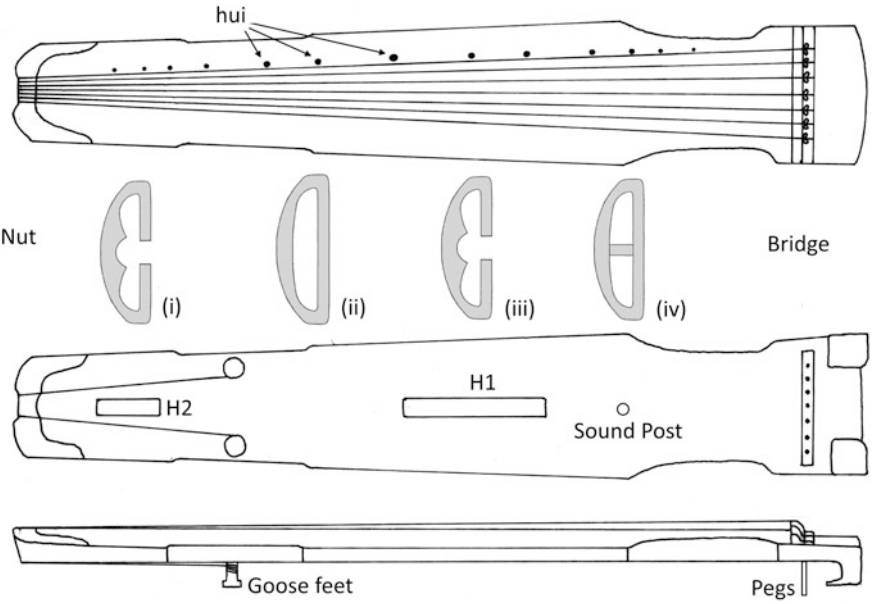


Fig. 2 General arrangement of a qin. Sound hole labels H1 and H2 refer to the dragon pool and phoenix pond respectively. The cross sections (*i*) and (*iii*) show how the inside of the front is carved to form the absorbers to restrict the hole openings; cross section (*ii*) shows a smooth top the cavity between the holes, and cross section (*iv*) shows a sound post

size. The thickness of the wood at the opening of the sound holes is approximately 10 mm. Opposite the hole on the inside of the front are raised areas ("absorbers") that restrict the path through the holes between the inside and outside to a thin gap about 20 mm wide. In the case of dragon pool, the raised part is also called a "tone receiver" (*na-yin* [12]). The cross sections (i) and (iii) in Fig. 2 show how the inside of the front is carved to form the absorbers to restrict the hole openings; cross section (ii) shows the unobstructed cavity between the holes, and cross section (iv) shows a sound post. Qins have one or two soundposts (sometimes none) placed on the central axis inside, a circular-sectioned "pillar of Heaven" between the dragon pool and the bridge, and a square-sectioned "pillar of Earth" between the two sound holes. Two of our examples have one soundpost (pillars of Heaven), the others have none. At the bridge end of the qin beyond the bridge itself is a small cavity called the "sound pool".

Data from three qins are shown in this chapter; they are labelled A, C and R and are all of recent (several decades old—rather than centuries) and standard construction, i.e. paulownia front, catalpa back. Qins A and R have been the subject of a recent paper [6]; qin A has relatively low frequency wood modes which couple weakly to the air modes, while qin R has greater wood/air coupling. Qin C is considered because it has a very similar modal structure to qin R but has silk-wrapped steel strings, whereas qin R has silk strings.

4 Playing the Qin

The qin is played in the manner of Fig. 3. The qin sits on a table resting on the "goose feet" and a soft pad placed to the left of the tuning pegs. The tuning pegs and "legs" hang over the edge of the table (the purpose of the "legs" is to prevent knocking the tuning pegs, not to support the instrument [11]). The strings, are tuned pentatonically (typically C2-D2-F2-G2-A2-C3-D3), but with much variation and often lower for purely silk strings. The literature on how to tune a qin properly (i.e. not with an electronic chromatic tuner) is extensive and varied, and beyond the scope of this chapter [12].

The qin is plucked with fingernails grown long for the purpose. There are three qualities of sound: open string (*san yin*, 散音), stopped string (*an yin*, 按音) and harmonics (*fan yin*, 泛音). For harmonics, a finger of the left hand is lightly touched against the string when it is plucked by a finger of the right hand. The positions where the string should be touched are marked by small circular inlays (often made of oyster shells) called *hui*.

The close proximity of the strings to the soundboard at the nut end allows for vibrato and portamento effects controlled by fingers of the player's left hand. The subsequent sliding sounds produced by the finger on the string are genuinely part of the qin's sound, and must be included in attempts to synthesize the characteristic sound of the qin [14].

Fig. 3 Modern qin being played by Vancouver musician Lin Min. The instrument on the *left* is known as a "banana-leaf" qin (Taylor Zhang photographs)

5 Construction

5.1 Wood

The paulownia and catalpa normally used in the qin are largely unknown in the Western instrument making. Paulownia is a hardwood that has unusual mechanical properties that lie between that of balsa (also a hardwood) and spruce (a softwood); it exhibits a very high Q, and is extensively used in the soundboards of many Asian instruments [15]. The catalpa and China fir used for the back of the qin are also unknown in the Western instrument making but have mechanical properties that are not unlike that of many softwoods known in the West.

The nomenclature of all woods commonly used for qin plates is given in Table 1. The potential for confusion caused by the logogram for "tong" is apparent, and is compounded by differing romanizations.

5.1.1 Paulownia

Paulownia tomentosa (Figs. 4 and 5) is native to parts of China, and is known in the West as the Empress or Princess tree, and in Japan as *kiri*. It can grow very quickly but the best tone woods come from trees growing slowly at altitude. The author of the Yu-ku-chai-ch'in-pu [11] states that the best wood grows on cliffs and has been burnt by lightning fire. If the lightning was accompanied by thunder, then the acoustic properties of the wood will be exceptional. Its mechanical properties, summarized in Table 2, are unusual, its density lying in between that of balsa and those of the lightest softwoods.

Table 1 Nomenclature of woods

Binomial	Chinese	Pinyin	Western
Paulownia tomentosa	泡桐	paotong, tong[a]	Princess tree, Empress tree
Firmiana simplex	梧桐	wutong, tong	Chinese parasol tree
Catalpa ovata	梓	zi	Chinese catalpa
Cunninghamia lanceolata	杉	shan	China fir

[a]"t'ung" in older texts using Wade-Giles romanization

Fig. 4 *Left* paulownia, quarter-sawn (*top*) and flat-sawn (*bottom*). *Right*: flat-sawn catalpa, part of an unfinished qin back plate (Author photographs)

Fig. 5 Paulownia trees in flower, Vancouver BC, 2016 (Author photographs)

In characterizing musical woods, Yoshikawa [15] plots the quantity cQ against ρ/c (where c is the speed of sound in the wood and ρ the density) and finds acoustic woods sit on one line with paulownia sitting at the extreme high-Q, low-ρ end. The surface of the wood is a little softer than spruce but can take a polish. On instruments like the pipa, the soundboard tends to suffer damage because it is not

Table 2 Mechanical properties of paulownia

Ref.	Density (kg/m^3)	E_L (GPa)	E_R (GPa)	E_T (GPa)	G_{LR} (GPa)	G_{LT} (GPa)	Q_L	Q_R	Q_T
[16]	317	4.3 ± 0.8							
[17]	260	7.3					170		
[18]	290	5.88	0.59	0.25					
[19]	308 ± 36	6.0 ± 1.0					139 ± 23		
[20]	277 ± 34	5.6 ± 1.6	0.63 ± 0.10	0.26 ± 0.04	0.52 ± 0.04	0.37 ± 0.05	140[a]	60[b]	60[c]
[4][d]	252	5.0	0.53	0.26	0.56	0.42			

[a]at 260 Hz; [b]at 730 Hz; [c]at 450 Hz

[d]Yang [4] also gives G_{RT} = 0.033 GPa, ν_{12} = 0.23 and ν_{13} = 0.49

varnished like the spruce of a violin. Under the heavy lacquer of a qin, the softness of the wood is not an issue.

5.1.2 *Firmiana simplex*

F. simplex (Chinese parasol tree) is a hardwood that has been used in the past for qin sound boards. The only available physical data are densities, and these range [21, 22] from 420 to 550 kg/m^3.

5.1.3 Catalpa

Catalpa trees bear a resemblance to paulownia trees, as both have heart-shaped leaves and similar pink flowers; however catalpas are distinguished by long seed pods. Numerical information on the mechanical properties of catalpa is not readily available; the two sources known to the present authors are given in Table 3.

5.1.4 China Fir

Like the situation with catalpa, numerical information is not readily available; the two sources known to the present authors are given in Table 4.

5.2 Lacquer

Frequently qins are covered with a thick (2.5 mm) paste made of lacquer and powdered deer horn that is polished (Fig. 6) and usually hides the wood grain beneath (p. 39 of Ref. [11]). Obataya et al. [25] have found that a thin layer of

Table 3 Mechanical properties of catalpa ovata

Ref.	Density (kg/m^3)	E_L (GPa)	E_R (GPa)	E_T (GPa)	ν_{12}	ν_{13}	ν_{23}	G_{LR} (GPa)	G_{LT} (GPa)	G_{RT} (GPa)
[4]	490	7.85	0.69	0.31	0.38	0.54	0.60	0.35	0.21	0.15
[23]	410	8.35								

Table 4 Mechanical properties of China fir

Ref.	Density (kg/m^3)	E_L (GPa)	E_R (GPa)	E_T (GPa)	ν_{12}	ν_{13}	ν_{23}	G_{LR} (GPa)	G_{LT} (GPa)	G_{RT} (GPa)
[4]	390	11.58	0.90	0.50	0.37	0.43	0.47	0.76	0.69	0.039
[24]		11 ± 2						0.65 ± 0.16		

Fig. 6 A high polish can be achieved on the lacquered surface of qin (Author photograph)

lacquer (up to 0.3 mm) has little effect on the longitudinal elastic constant (E_L) of 3 mm thick spruce, but may increase the radial value (E_R) by up to 50 %. Lacquer also markedly increases the damping, particularly in the radial direction. Thus it may be expected that applying a much thicker layer to an uncoated qin, even with plates 10 mm thick, will have a measureable effect on its vibrational properties. The authors are not aware of any acoustical work on the lacquering of qins.

5.3 Strings

Historically, qin strings were made of silk, which are expensive and fragile. Around 1966 the production of high quality silk strings was curtailed and wrapped-steel strings became dominant. Steel strings generally sound louder and have other characteristics that silk strings don't have. The historic "warm" sound character of the qin was due to the silk strings, but some players have come to prefer the "metal sound", which to others can be much like the sound of an electric guitar. When he lived, the renowned qin player, Wu Zhaoji (1908–97) believed that metal strings should only be used when one has no choice. He used a *pianfeng* style (striking the steel string at an angle so that the right half rather than the middle of one's fingertip touches the string) to ameliorate the metallic sound [26]. Starting around the turn of this century there has been a resurgence in the production of higher quality silk strings and concerted efforts to produce strings using the traditional methods described in ancient qin handbooks [11], despite to loss of prime silkworm habitats. Only the highest quality strings can be set to the same pitch as metal strings without frequent loss due to breakage.

Silk strings usually consist of multiple twisted cords, much like rope, which are glued together to form a string. The manner of production is such that some of the thinner strings become a core for the three lower pitch strings; to these cores are

added a wrapping to augment the mass of the string. A set of silk strings usually has eight strings, one of which is an extra string 7 (the thinnest) as these are most likely to break.

Liang and Su [27] have analyzed the tonal qualities of nylon-wrapped silk-wound steel qin strings mounted on a rigid frame, but not when mounted on an instrument. They found that these strings had a softer spectrum than even nylon-wound steel pipa strings and much softer timbre than steel-wound steel guitar strings. Tse [28] has examined the contribution of longitudinal vibration modes (that occur between about 1.1 and 1.4 kHz for silk strings, 1.2–1.7 kHz for silk-wound steel) to the sound of a qin. Longitudinal modes are particularly prominent in the sound of a finger sliding on a string [14] when playing portamento.

The silk strings are made of individual strands, whose traditional number [7] is given in Table 5. These numbers are very closely proportional to the inverse of the playing frequency, $1/f$. At first sight this seems odd, as an ideally uniform tension should dictate that the cross-sectional area of the string should be proportional to $1/f^2$. However inspection reveals that the angle of the windings varies greatly from string to string [12] (Fig. 7), with the outer part of the low strings being tightly wrapped to increase the linear density while keeping the stiffness low. In addition, Yang's measurements of seven sets of new silk strings [5] shows that, although there is large variation in tension between sets, there is a generally rising trend from string 1–7. The physical properties of the silk strings on our qin R are shown in Table 5.

The strings of a qin are knotted at the bridge to a thicker silk thread which passes through a hole to the tuning pegs on the underside of the sound box. When a peg is turned, the thicker thread tied to it gets twisted and the tension on it increases, which in turn raises the tension of the strings. This tension holds the top face of the peg against the soundbox, and friction prevents it from moving [12]. This tuning

Table 5 Silk string dimensions and tensions, taken from qin R, tuned to A1-B1-D2-E2-F3#-A2-B2

String	Fundamental (Hz)[a]	Diameter (mm)[a]	Tension (N)[a]	Tension (N)[b]	Silk strands[c]
1	55	1.77	45	29–77	108
2	62	1.64	50	32–68	96
3	73	1.38	46	33–73	81
4	82	1.25	51	45–76	72
5	93	1.10	55	44–84	64
6	110	0.99	61	54–85	54
7	123	0.79	45	53–88	48

The density and elastic constant of the strings were found to be 1200 kg/m^3 and 12.0 GPa respectively

[a]qin R; [b]from Yang [4]; [c]nominal, from Lawergren [7]

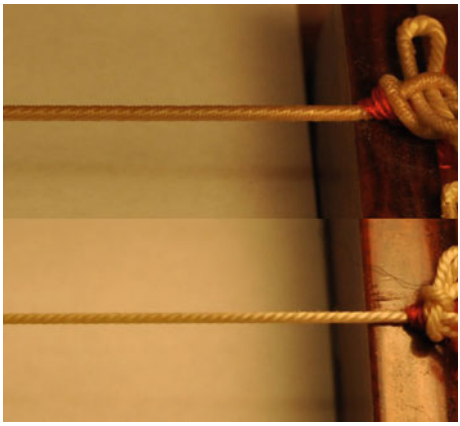


Fig. 7 Silk strings 1 and 6 on qin R, showing the differing angle of windings (Author photograph)

mechanism is unique to the qin, and allows the pegs to be arranged in one compact row largely out of sight underneath the instrument.

The 60 dB decay times for qin silk strings were measured to be of the order 20 s for the lowest and 10 s for the highest open strings.

6 Vibroacoustics

6.1 Acoustics of Long Soundboxes

The acoustic character of an instrument is partly determined by the nature of the strings and how they are excited, and partly determined by the vibrational behaviour of the soundbox. The radiation from a sound box is determined by its surface velocities and air velocities from the sound holes (if they exist). The surface velocities result from a force being applied by the strings, in this case to the bridge and also at the nut. The velocities of the wood surfaces are straightforward to measure, with an impact hammer and small accelerometer, for example. Finding the velocities at the opening of the sound holes is more involved, requiring a small array of microphones [6]. The air velocity field at the sound hole is controlled by the cavity modes of the box (and how these mix with the wood modes, to be precise). For long soundboxes, like those of the qin and of the Western harp, are quasi-one-dimensional modes; there is no real "A0" mode (often referred to colloquially as the Helmholtz mode) of the type seen in more compact soundboxes like those of the guitar or violin [29], i.e. where all the air in cavity is moving in phase. The cavity modes of the qin have been modelled by the one-dimensional transfer matrix method adapted from the study of woodwind instruments [6].

6.2 *Surface Velocities*

6.2.1 Wood

For accelerometer-based measurements the instruments were suspended on two bungees at the places where a qin being played would be supported on a table, i.e. at the goose feet and just ahead of the bridge. The support positions are close to the nodes of the first bending mode [6], which is probably not a coincidence.

The lowest five modes of qin A are shown in Fig. 8. Qin A is chosen to illustrate the "wood modes" because there is little interaction with the cavity modes in this

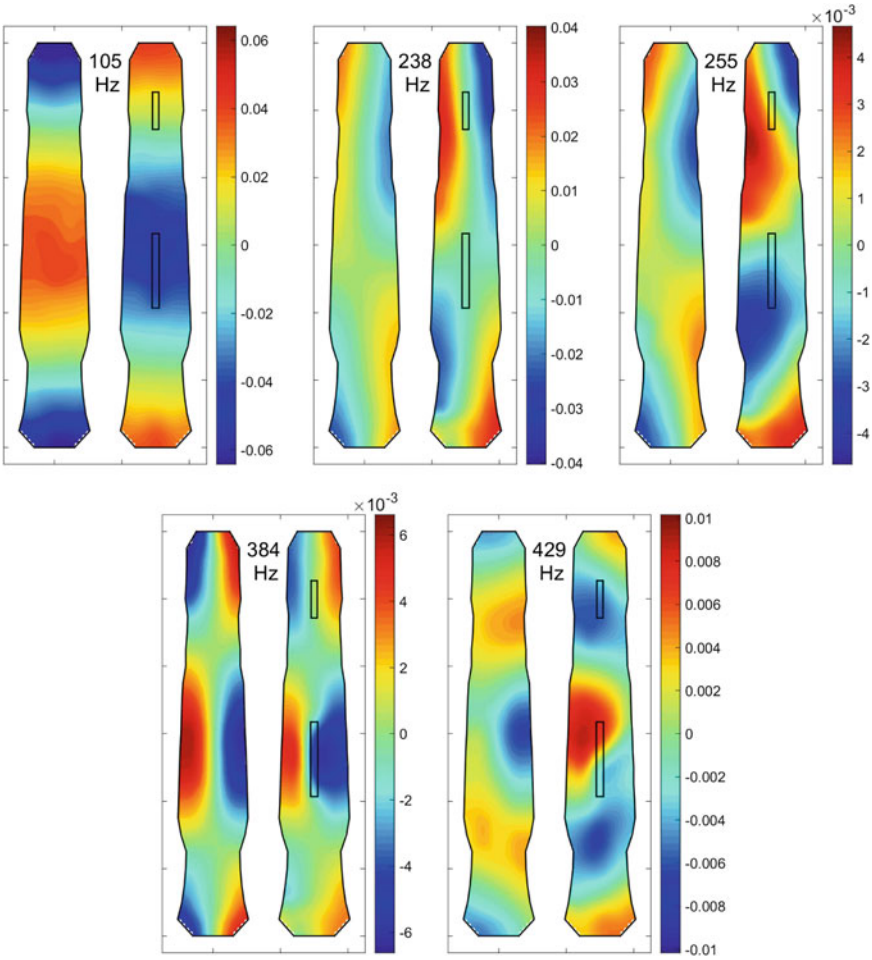


Fig. 8 The first five wood modes of qin A: first bending mode (0, 2) at 105 Hz; first torsional mode (1, 1) at 238 Hz; second bending mode (0, 3) at 255 Hz, mode (2, 1) at 384 Hz; mode (4, 0) at 429 Hz. Air velocities in the holes are not shown

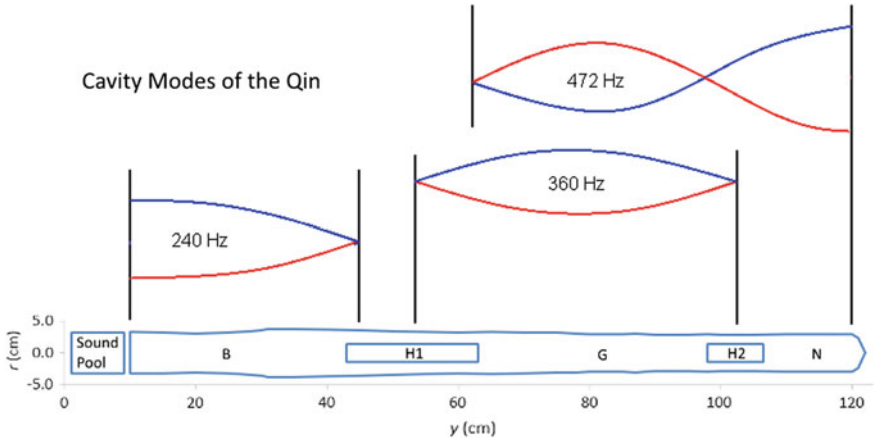


Fig. 9 Cavity structure of the qin, expressed as a cylindrical tube with varying radius and the same cross-sectional area profile as the qin. The lowest three cavity modes are shown. The frequencies given are those observed in qin A when immobilized with sandbags

instrument. All the modes can be described in terms of those of a flat plate, i.e. the front and the back move more-or-less in unison, and only start to go their separate ways well above 500 Hz.

6.2.2 Air

The air modes in the qin cavities were measured using a linear array of small electret microphones mounted on a thin flexible rod and inserted into the cavity through the sound holes. The cavity was excited by a small loudspeaker external to the sound holes. The qin was either suspended free to vibrate, or encased in many small sandbags to immobilize the wood; in this way the magnitude of the wood-air coupling could be assessed. The basic modal structure is given in Fig. 9 with frequencies given for qin A. The values of the frequencies were dependent on details of how the inside of the qin was carved and could vary by tens of Hz [6]. In the case of qin A, the cavity modes were distant enough from the bending modes not to provoke any couplings. The longitudinal nature of the cavity prevented coupling with the torsional modes, some of which did lie near cavity frequencies.

6.2.3 Wood and Air

To measure the radiativity each qin was suspended, at the centre of a circular 30-microphone array of approximately 1 m radius, inside an anechoic chamber. Radiativities (R), defined here to be the mean sound pressure level produced by a force of 1 N (rms) applied at a given frequency to the bridge, are reported in dB re Pa/N.

The low frequency structure of qin R (Fig. 10) is somewhat more complicated than that of qin A which has been previously discussed. The lowest bending mode (0, 2) for qin R lies at 156 Hz, i.e. above the fundamental frequencies of all the strings. The lowest cavity mode (from cavity section B, weakly excited at the bridge) is just visible in H1 at 192 Hz, but the wood velocity is too small to discern a shape. The (1, 1) torsion mode is split into 296 and 310 Hz by a cavity mode. The second bending mode (0, 3) is hard to identify, and may be mixed with the predominantly cavity modes at 389 and 426 Hz. The third bending mode (0, 4) (not shown) is split into modes at 533 and 593 Hz by the motion of the back and the cavity air.

Deflection shapes of the first five wood modes of qin A are shown in Fig. 8. The low modes of this qin are largely unaffected by cavity modes and can be given designations (n_x, n_y) denoting the number of nodelines in the x (across the width) and y (along the length) directions. The first bending mode (0, 2) is at 105 Hz, the first torsional mode (1, 1) at 238 Hz; the second bending mode (0, 3) at 255 Hz, mode (1, 2) at 384 Hz and mode (0, 4) at 429 Hz. The Q-factors for the low bending modes are high (~ 100) when the qin is suspended on bungees, but much lower when placed on soft pads on a table [6].

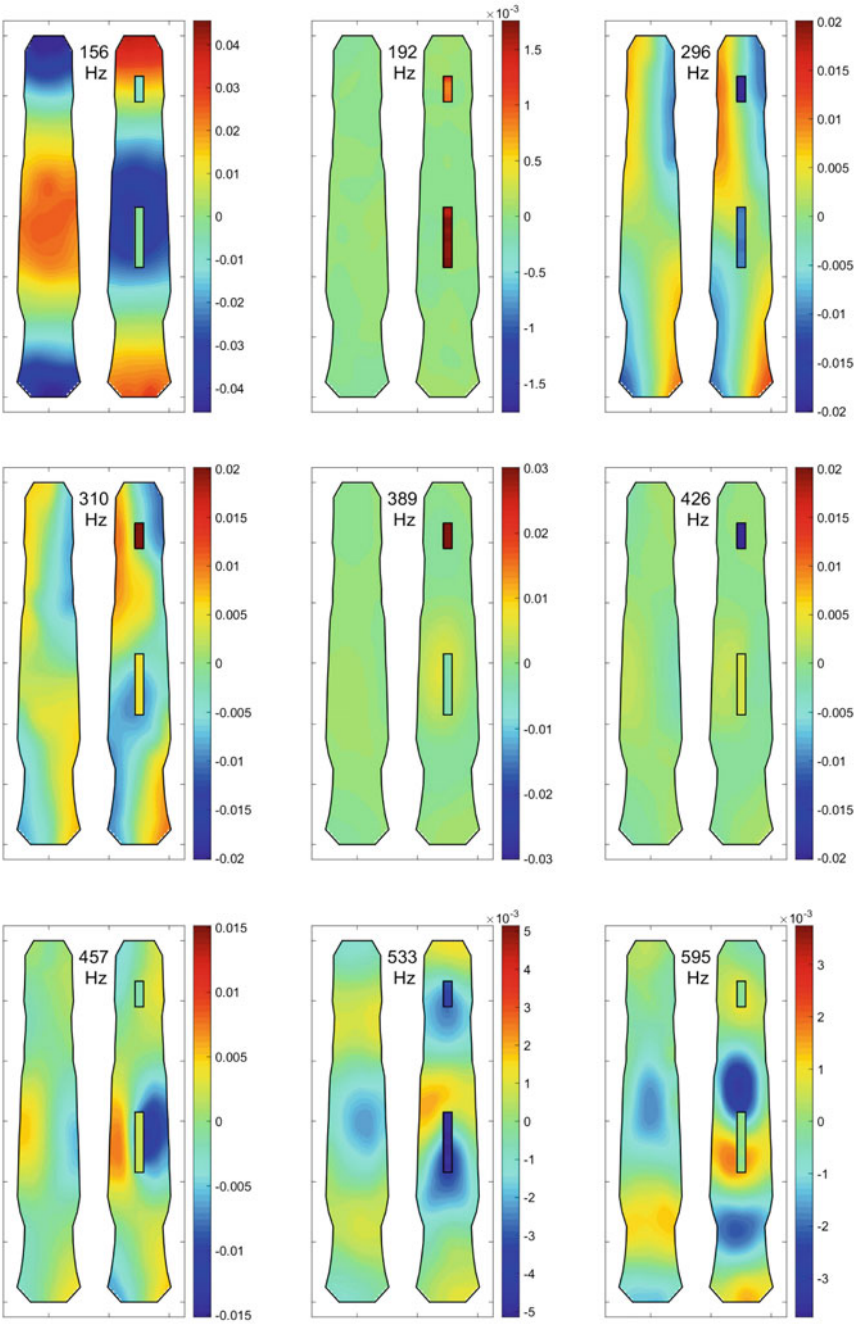
The frequencies of the first three bending modes of the five qins in Ref. [6] have a mean spacing of 1:2.37:3.75, i.e. closer to 1:2:3 than the 1:2.76:5.40 ratio expected for a uniform free beam. A similar feature has been noted for the koto [30, 31].

7 Sound

The sound of a qin is determined not only by the quality of the strings and the soundbox, but also the table upon which it sits and the room in which it is played. In this section we compare measurements made upon a qin suspended in an anechoic chamber, and those made with the qin sitting on trestles and a table in a small room, noting that the qin, historically, has been played in small spaces.

The typical structure of the qin's radioactivity spectrum is that the first bending modes are prominent with little modal overlap, and that higher bending and torsional modes are split by the air in the cavity. At higher frequencies the sequence of bending modes which must be present merges with a denser spectrum of cavity modes and the bending modes lose their distinct identity. The qin radiativity curves appear to be similar in magnitude and general shape to that of the simple beam, but the density of modes at higher frequencies tend to smooth out the curve [6].

Most zithers familiar in the West have sound holes on the front surface, operating in open air. However the qin in its conventional playing configuration has its sound holes close to and facing the hard reflecting surface of the table upon which it


156 Hz
0.04
0.03
0.02
0.01
0
-0.01
-0.02
-0.03
-0.04
192 Hz
×10^-3
1.5
1
0.5
0
-0.5
-1
-1.5
296 Hz
0.02
0.015
0.01
0.005
0
-0.005
-0.01
-0.015
-0.02
310 Hz
0.02
0.015
0.01
0.005
0
-0.005
-0.01
-0.015
-0.02
389 Hz
0.03
0.02
0.01
0
-0.01
-0.02
-0.03
426 Hz
0.02
0.015
0.01
0.005
0
-0.005
-0.01
-0.015
-0.02
457 Hz
0.015
0.01
0.005
0
-0.005
-0.01
-0.015
533 Hz
×10^-3
5
4
3
2
1
0
-1
-2
-3
-4
-5
595 Hz
×10^-3
3
2
1
0
-1
-2
-3

◀ **Fig. 10** Interaction of wood and air modes. Wood surface velocities and sound hole air velocities of qin R, expressed in m/s per newton of force applied to the treble side of the bridge. The (0, 2) mode is at 156 Hz; the lowest cavity mode is at 192 Hz; the (1, 1) mode is split between 296 and 310 Hz; the pair at 389 and 426 Hz have characteristics of a split (0,3) mode. The (1,2) mode lies at 457 Hz; above that many of the modes start to exhibit independent behaviour in the front and back motions, e.g. at 533 and 595 Hz

sits. The acoustic implications have not been thoroughly investigated, although preliminary sound radiation measurements with and without a table are reported below. In addition to reflection, it may be imagined that the qin/table combination creates an addition air volume which has its own resonant characteristics.

7.1 Table Effects

According to the Mei-an ch'in-p'u [12] the tables can be fixed and made of earthenware bricks and a wooden frame, or, more commonly, be a movable one 2′ 2″ (Chinese measure: 733 mm) high, made of light woods like paulownia or pine. The choice of light, acoustic woods will undoubtably add some acoustic character of their own to the qin's sound. However, while the table length and width are specified, the thickness is not, so the effect is unlikely to be uniform. Observation of present-day performance indicates that the tables are not chosen to a fixed standard.

Figure 11 shows the differences in radiated sound when the qin is either suspended in an anechoic chamber or placed in conditions more like those of a real performance, on trestles (i.e. open underneath) or on a more conventional table. It is plain that certain modes selectively radiate in different conditions. For example, the lowest air mode of qin R at 192 Hz shows up the clearest when the qin is on a table. Conversely, the lowest bending mode at 156 Hz only shows when the qin is suspended close to the mode's nodal points. More significantly, the air modes around 500–600 Hz are more prominent when the qin is on a table, and this may be the largest contribution to the audible difference in the sound quality of the instrument.

The sound radiation from a plucked qin was measured inside a small room. The qin was placed either on a table, on soft pads, as per the normal manner of playing, or on two trestles. A comparison of the sound radiation in the two situations was used to investigate the effect of having a reflecting surface close to the sound holes of the instrument. Two qins were chosen with closely matched lowest bending modes ("qin C" at 158 Hz and "qin R" at 156 Hz); qin C had strings of nylon-wrapped steel, tuned as per Table 6, and qin R has silk strings, tuned a semitone and a half lower as per Table 5. The strings were plucked with the fleshy part of a finger at the last hui (1/8 of the length of the string from the bridge). The microphone was placed 20 cm above the centre of the instrument.

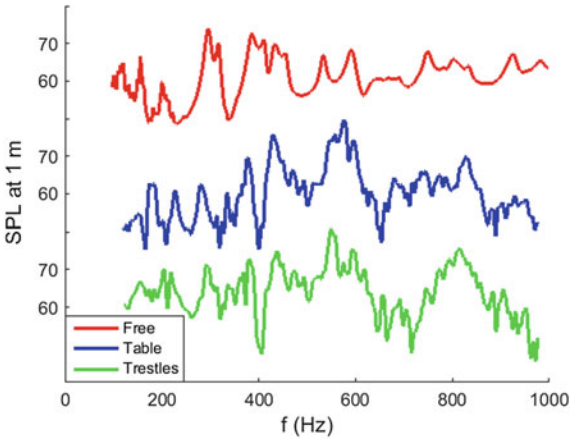


Fig. 11 Spectra of sound radiated by qin R in different conditions. "Free" indicates hanging on bungees in an anechoic chamber; "table" indicates sitting on a table in a small room supported by soft pads at the bridge end and under the goose feet, in the manner of a normal performance; "trestle" is the same as "table" except that the qin is supported by two open trestles at the bridge end and under the goose feet

Table 6 Nylon-wrapped steel string dimensions and tensions, taken from qin C, tuned to C2-D2-F2-G2-C3-D3

String	Outer diameter (mm)[a]	Core diameter (mm)[a]	Tension (N)[a]	Tension (N)[b]
1	1.42	0.69	90	72–85
2	1.25	0.64	93	80–90
3	1.11	0.40	69	60–65
4	0.98	0.37	69	65–70
5	0.92	0.28	68	67–74
6	0.82	0.30	85	55–74
7	0.64	0.28	76	66–76

The anharmonicities were found to be negligibly small
[a]qin R; [b]from Yang [5]

The effects of the table are not obvious in Fig. 12. However, there seems to be some evidence that the table enhances the region from 500–700 Hz, which is probably related to the similar feature seen in the tap spectra of Fig. 11.

8 Simulations

We report here two finite element method (FEM) simulations of the qin, one using the Abaqus [32] software reported in Yang's Ph.D thesis [4, 5] with no fluid-structure coupling, and one using Comsol [33] Multiphysics 5.2 with fluid-structure coupling [33] which is currently being attempted by one of the present authors (KC). The following sections will deal with general issues relating to constructing a mathematical model of the qin, and will be followed by a brief summary of results so far.

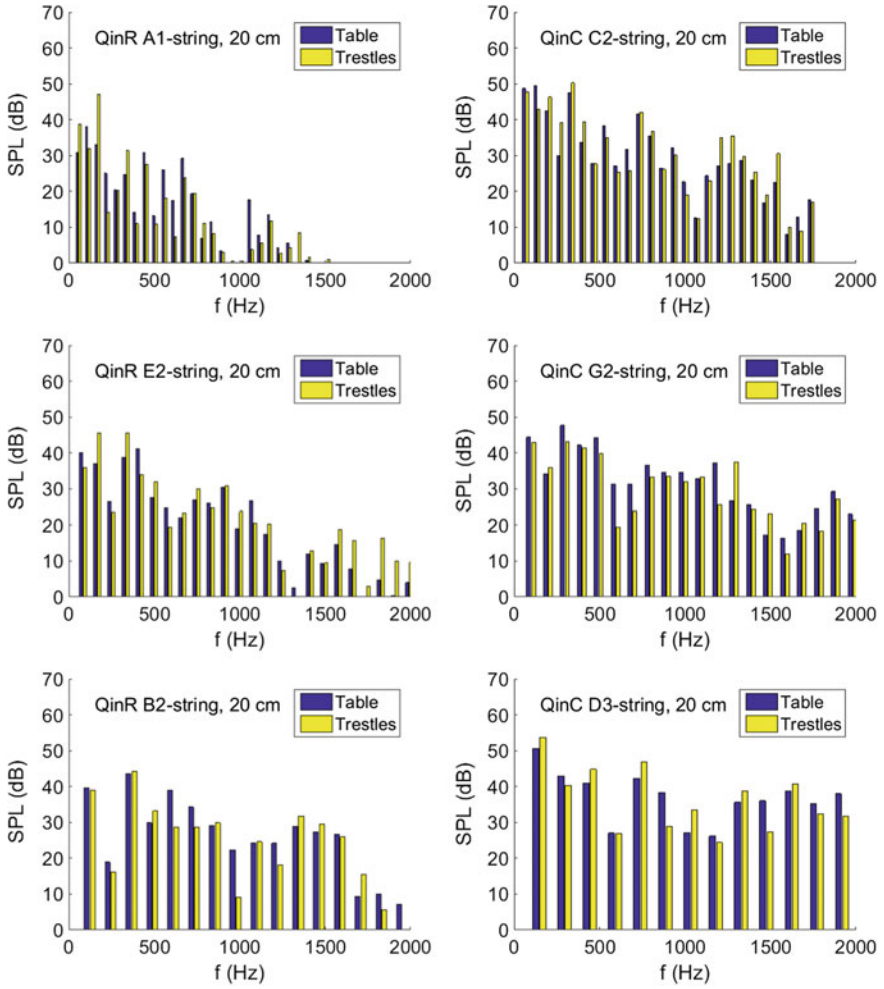


Fig. 12 Strength of partials in the sound of plucked open qin strings. *Left* silk strings. *Right* silk-wrapped steel strings

8.1 Construction of the Finite Element Model of the Qin

The qin is a highly crafted instrument with an "organic" shape that has many curves and much ornamentation; it does not have a simple geometric shape that lends itself easily to the construction of a computer model for simulation. There are a number of ways to construct "organic" shapes for computer model. In the Comsol project the use of non-linear rational B-splines ("nurbs") proved to be helpful. More

significantly, the modelling package in Comsol allowed the construction of the instrument as a series of cross-sections that could be "lofted" together into a smooth arbitrary shape. The concept of lofting is most frequently associated with ship-building in which cross-sections are joined to form smooth curves and an elegant shape. There is, however, one limitation: the lofted shape must have continuous tangents at all points along the curve but good approximations at corners still need to be made. The loft structure is shown in Fig. 13.

A solid object is then formed in the Comsol modelling package as seen in Fig. 14. The next step is to take this solid object and repeating the process to form a duplicate, whose size is reduced in each dimension to match the measured instrument walls. This process allows Boolean subtraction of the two shapes in the modelling package to yield a hollow shape representing the front shell of the instrument without a back plate, the two pieces being made from different materials. The different physical properties need to be mapped onto the correct shape in the process of completing the model for accurate simulation of the acoustic properties of each wood. The final step is to construct the back plate in a similar manner, with sound holes punched into the base (Fig. 15). The complete model is imported into Comsol 5.2 with the Acoustics module for the next step of setting up the finite element model.

The model can then be placed into a sphere of air and meshing applied to generate the finite elements (Fig. 16). This is appropriate when the interaction with air is being considered, such as for scans, frequency response studies or the simulation of a note. For simple eigenmode studies enclosing in air serves no purpose. In this study 42,973 elements were generated, leading to the need to solve for 647,443 degrees of freedom (dof).

Paulownia and catalpa are strongly anisotropic, and the FEM model requires the full set of elastic constants. This is often problematic due to the natural variability of the wood and the limited information available. It was thus prudent to start with only one wood, catalpa, from which the back plate was made to test the model. The properties of paulownia for the top plate were then added.

8.2 Back Plate Study

The Comsol FEM model was tested with a qin back plate (Fig. 15) constructed by Jim Binkley from a flat-sawn piece of catalpa ovata, 13 mm thick. The mode shapes and frequencies were measured by standard accelerometry. Several longitudinal bending modes (0, n) and torsional modes (1, n) were observed; no transverse bending modes (m, 0) were seen. Thus the measurements were most sensitive to E_L and G_{LT}. The model reproduced the lowest ten modes well (see Table 7; Fig. 17).

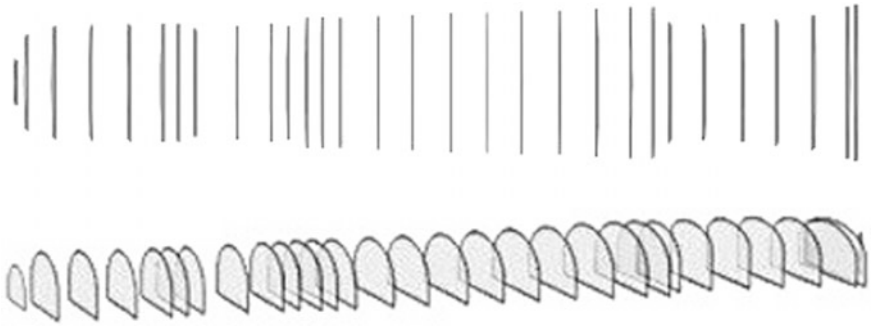


Fig. 13 The lofts that form the qin shape prior to application into the model

Fig. 14 A solid qin model as created in Comsol

Fig. 15 Creating the back shell for the model

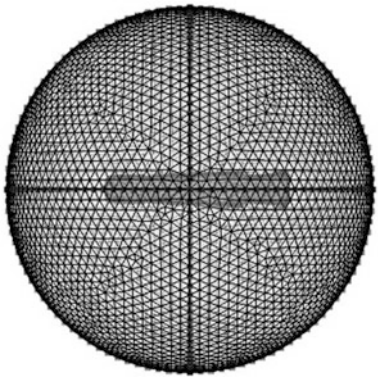


Fig. 16 The model of the qin in a sphere of air with meshing applied

8.3 FEM Analysis of Historical Qins

Yang [4, 5] took geometrical data from six historical qins, in particular the famous Tang Dynasty "Jiu Xiao Huan Pei" instrument in the Palace Museum in Beijing. He used these data to create FEM models of the instruments and predict the frequencies

Table 7 Modal frequencies of catalpa back plate—a comparison of experimental values and FEM simulation

Mode	Experiment frequency (Hz)	Model frequency (Hz)
(0, 2) Longitudinal	37	36
(0, 3) Longitudinal	97	95
(0, 4) Longitudinal	190	185
(0,5) Longitudinal	311	312
(0, 6) Longitudinal	460	467
(1, 1) Torsional	83	88
(1, 2) Torsional	164	163
(1, 3) Torsional	255	250
(1, 4) Torsional	372	363

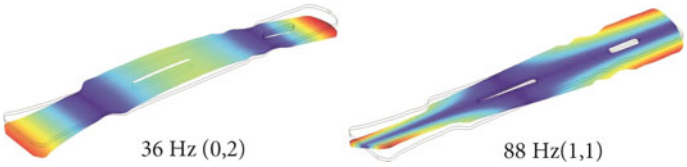


Fig. 17 Back plate eigenmodes identified using the finite element model. The lowest bending and torsional modes of the back plate from the finite element model

and shapes of the wood modes. He did not include fluid-structure coupling and was not able to compare the simulations with data from these museum instruments. Because it is often not possible to identify the woods used in lacquered instruments, nor even to ascertain the masses, Yang ran simulations for all credible combinations of woods used in the front and back plates: paulownia-fir, paulownia-catalpa (two varieties of the latter), paulownia-paulownia and fir-fir. He also looked at the variation of mode frequencies with plate thicknesses, and at the effect of the qin's soundpost.

Although it is hard to make comparisons between FEM results for historical qins and vibrational data from modern versions, several features stand out.

1. The bending modes given by the FEM are lower than those seen in modern instruments. The first and second bending modes assuming paulownia-catalpa ovata construction for the six instruments were 74–109 Hz (the lowest value being for a banana-leaf type) and 168–193 Hz respectively. The five modern instruments measured in Ref. [6] showed values for the first and second bending modes of 105–156 Hz and 235–310 Hz.
2. The FEM indicated the existence of a mode lower in frequency than the first bending mode (at 68–106 Hz), one that was essentially an isolated mode of the back plate; this did not appear in the modern instrument data. There were other similar modes in between the bending modes that did not appear in the data.

3. The FEM showed sideways modes starting at 400 Hz that may be important but as yet there are no modern instrument data for comparison.
4. The FEM indicated that the effect of the existence or not of a soundpost on the lower bending mode frequencies was very small, of the order of 1 Hz or less. For complex modes with different front and back shapes, the effect was often larger, up to several %, as the soundpost stiffened the qin body.

The thesis also contains analyses of sound spectra with different string types and calculations of modes with somewhat thicker front plates.

9 Concluding Remarks

There are few reports in the acoustical literature regarding any plucked Chinese string instruments. Shih-yu Feng's brief 30-year-old article on the pipa [34] notes that the radiation from this instrument is strongest in the 400–600 Hz region. Yoshikawa [35] has made measurements on a Japanese relative of the pipa, the biwa, and made similar observations about the radiation. In particular, he notes that the choice of woods and construction of the biwa seem to aim at enhancing the higher harmonics produced by the sawari mechanism of the biwa's nut and frets. Two of the present authors have measured pipas and yueqins [36] and also concluded that their radiation also favoured higher harmonics over the string fundamentals. Whether this feature is true for the qin is not clear. The qin string fundamentals are very low, mostly lower than the lowest vibration mode of the sound box. Nonetheless, the fundamentals of all strings are clearly visible in the radiated sound spectra, typically as prominent as the most prominent higher partials, although there is some enhancement by the higher air modes around 400 Hz.

It is plain that we are a long way from any prescription for making a "good" qin, of the type that exists for the violin and guitar. In the case of these two Western instruments, it is known from examining old, successful instruments how to craft the front and back plates in such a way as to have a chance of producing a competent final product [37]. Even so, an understanding of how the behaviour of the plates influence that of the soundbox is only just beginning to emerge, as fluid-structure coupled FEM simulations become more reliable [38]. The next calculational step is to bring the full fluid-structure coupled FEM model into some agreement with the observed modes of a real instrument. At the same time, one of the current authors (CW) is dismantling a qin with the intention of measuring the vibrational behaviour of its components and reassembling it in a manner such that the absorber and soundpost can be attached or removed, thus providing confirmational data for the FEM model.

The physical nature and origin of the sound of a qin is a big subject, as big as that for the violin family, which has occupied a fair fraction of the world's musical acoustics effort for the last two centuries.

Acknowledgments The authors thank Lin Min, Taylor Zhang, Alan Thrasher and the UBC Chinese Orchestra for the loan of instruments, Jim Binkley for the gift of a catalpa back plate and illuminating discussions on making and playing qins, and Dan Blickenstaff of the American Paulownia Association for supplying wood samples.

References

1. Liang, M.-Y., & Lam, J. S. C.: Qin. In: Grove music online. Oxford http://www.oxfordmusiconline.com. Cited 16 Mar 2016
2. Sachs, C.: The Rise of Music in the Ancient World: East and West, p. 108. Norton, New York (1943)
3. van Gulik, R.H.: The Lore of the Chinese Lute. Sophia University, Tokyo (1940)
4. Yang, F. 杨帆: 对不同材质古琴 琴体声学特性的有限元分析 (Dui butong caizhi guqin qin ti shengxue texing de youxian yuan fenxi; Guqin of different materials: finite element analysis of the acoustic characteristics of the qin body). In: Chinese. Musical instrument magazine pp. 14–20 (2014)
5. Yang, F.: 杨帆: 古琴震动体与共鸣体声学特性研究 (Guqin zhendongti yu gongmingti shengxue texing yanjiu; Research into the vibration and acoustic characteristics of the guqin). In Chinese. Central Conservatory of Music Press, Beijing (2015)
6. Waltham, C.E., Koster, E., Lan, Y.: An acoustical study of the qin. J. Acoust. Soc. Am. **139**, 1592–1600 (2016)
7. Lawergren, B.: Strings. In: Music in the Age of Confucius, pp. 65–85. University of Washington Press, Seattle (2000)
8. Furniss, I.: Music in ancient China, pp. 67–71. Cambria, Amherst (2008)
9. Zheng, M. Z.: Palace Qin, pp. 47–63. Forbidden City Museum, Beijing (1991)
10. Toshiroo, K.: Reconstructed music instruments of ancient East Asia, pp. 48–57. Ongaku No Tomo Sha (1994)
11. Binkley, J.: Abiding with Antiquity: Translations from the Yu-ku-chai-ch'in'pu, pp. 11–73. Lulu Press. Also available online at http://web.cecs.pdx.edu/jrb/chin/index.html. Cited 25 June 2015 (2006)
12. Lieberman, F.: A Chinese zither tutor, the Mei-an Ch'in-P'u, pp. 8–16. University of Washington Press, Seattle (1983)
13. Wu, Z. 吴钊: 绝世清音 (Jue shi qing yin; Very clear sound of the epoch). In Chinese. pp. 9–37. Gu Wu Xuan, Suzhou (2005)
14. Penttinen, H., Pakarinen, J., Välimäki, V., Laurson, M., Li, H., Leman, M.: Model-based sound synthesis of the guqin. J. Acoust. Soc. Am. **120**, 4052–4063 (2006)
15. Yoshikawa, S.: Acoustical classification of woods for string instruments. J. Acoust. Soc. Am. **122**, 568–573 (2007)
16. Akyildiz, M.H., Kol, H.S.: Some technological properties and uses of paulownia (Paulownia tomentosa Steud.) wood. J. Environ. Biol. **31**, 351–355 (2010)
17. Aizawa, H.: Frequency dependence of vibration properties of wood in the longitudinal direction, M.Sc. thesis, Faculty of Engineering, Kyoto University (in Japanese, quoted in Ref. [35])
18. Nakano, T.: Coarse graining of wood cell arrangement and density of dependence of elasticity. Holzforschung **67**, 67–73 (2013)
19. Obataya, E., private communication (2014)
20. Waltham, C.: Report for the American Paulownia Association (2016, unpublished)
21. Firmiana simplex. In: encyclopedia of life. http://eol.org/pages/584813/overview. Cited 12 Apr 2016
22. Firmiana simplex. In: Xycol database http://xycol.net. Cited 12 Apr 2016
23. Meier, E.: Wood database. http://www.wood-database.com. Cited 6 Apr 2016

24. Cho, C.-L.: Comparison of three methods for determining young's modulus of wood. Taiwan J. Sci. **22**, 297–306 (2007)
25. Obataya, E.: Effects of oriental lacquer (urushi) coating on the vibrational properties of wood used for the soundboards of musical instruments. Acoust. Sci. Tech. **22**, 27–34 (2000)
26. Wong, S.-C.: 黄树志: 从琴弦探讨古琴过去、现在与未来的发展路向 (Cong qin xian tantao guqin de guoqu. xianzai yu weilai de fazhan luxiang; Through qin strings, inquiring about the guqin's past, present and future path of development, http://www.silkqin.com/03qobj/strings/shuchee1.htm. Cited 13 Apr 2016
27. Liang, S.-F., Su, A.W.Y.: Modeling and analysis of acoustic musical strings using Kelly-Lochbaum lattice networks. J. Inf. Sci. Eng. **20**, 1161–1182 (2004)
28. Tse, C.-Y.: 谢俊仁: 古琴音色与琴弦之纵向震动 (Guqin yinse yu qin xian zhi zong xiang zhendong; Sound quality of the qin and longitudinal vibration of its string). Chinese. Qixianqin Yinyue **5**, 29–30 (1999)
29. Bell, A.J.: The Helmholtz resonance and higher air modes of the harp soundbox. Catgut Acoust. Soc. J. **3**, 2–8 (1997)
30. Ando, Y.: Acoustics of sohs (kotos). Proc. 12th Intl. Congr. Acoust., **3**, K1–K5, Toronto (1986)
31. Coaldrake, K.: Extending the musical understanding of timbre in the Japanese Koto through COMSOL multiphysics simulation, COMSOL 2012 (Tokyo, 2012). http://www.comsol.com/paper/download/159347/coaldrake_poster.pdf. Cited 19 Mar 2015
32. Abaqus, http://www.3ds.com Dassault Systèmes, 10 rue Marcel Dassault, 78946 Vélizy-Villacoublay, France
33. Comsol Inc., http://www.comsol.com 100 District Avenue, Burlington, MA 01803, USA
34. Feng, S.-Y.: Some acoustical measurements on the Chinese musical instrument P'i-P'a. J. Acoust. Soc. Am. **127**, 3203–3211 (1984)
35. Yoshikawa, S.: A comparison of string instruments based on wood properties: biwa vs. cello. Acoust. Sci. Technol. **29**, 41–50 (2008)
36. Waltham, C.E., Koster, E., Kotlicki, A., Simard J., Wolfe, N: Acoustic radiation from the Pipa and Yueqin, Proc. Meet. Acoust. **035004** (2013)
37. Fletcher, N.H., Rossing, T.R.: The physics of musical instruments. Springer, New York (1991)
38. Gough, C.E.: A violin shell model: vibrational modes and acoustics. J. Acoust. Soc. Am. **137**, 1210–1225 (2015)

Authors Biography

Chris Waltham Before musical acoustics, Chris Waltham spent twenty years working on the Sudbury Neutrino Observatory project. In the last dozen years he has worked on the acoustics of string instrument soundboxes, specializing in harps and qins. He is an amateur luthier, and has made several harps and violins, playing one of the latter in a Vancouver community orchestra.

Kimi Coaldrake is Associate Professor in the Elder Conservatorium of Music at the University of Adelaide, Australia. She is a Fulbright Scholar and Affiliate-in-Research at the Reischauer Institute of Japanese Studies at Harvard University. She is also a professional performer on Japanese koto (zither). She has published in the areas of Japanese music theatre and the incorporation of tradition into Japanese contemporary culture. Her current research uses Finite Element Analysis to investigate the acoustic properties of the koto and its relationship to the culture of sound in Japan.

Evert Koster is a sessional lecturer at the University of British Columbia Department of Physics and Astronomy. His research background is in NMR studies. He has been playing and studying qin for about 20 years. Other interests include Chinese paintings and martial arts, as well as Chinese food.

Yang Lan (born in Chongqing, China) gained his MSc with Chris Waltham working in musical acoustics. Yang is an amateur player and maker of the Chinese end-blown flute, the xiao. He has done measurement, modeling and optimization of the xiao and participated in the study of qin. Yang is now a PhD student in UBC/TRIUMF studying neutrino-less double-beta decay.

Tone Production of the Wurlitzer and Rhodes E-Pianos

Florian Pfeifle and Malte Münster

Abstract Two idiomatic examples of electro-acoustical keyboards played since the 60s to the present day are the Wurlitzer E-Piano and the Rhodes E-Piano. They are used in such diverse musical genres as Jazz, Funk, Fusion or Pop as well as in modern Electronic and Dance music. Their unique sound, that is comparable on a generic level, shows distinctive varieties in timbre and decay characteristics. This can be attributed to their specific mechanical-electromagnetic/electrostatic tone production. In this treatise, a description and comparison of the tone production mechanisms are presented based on measurements taken on both instruments, a Rhodes Mark II and a Wurlitzer EP300. The measurements include high-speed camera measurement and tracking of the primary mechanical sound production mechanisms as well as audio recordings of the unamplified instrument signal. It is highlighted that the different timbre can be attributed to different characteristics of the pickup systems of both instruments. In the case of the Rhodes, characteristic sound properties emerge due to the interaction of the mechanical motion of a small tine interacting with the magnetic field (H-field) of the pickup. In the case of the Wurlitzer a vibrating steel reed acts as the zero potential electrode of a capacitor inducing an alternating current due to changes in the electro-static field (E-field). The measurements are compared to a FEM model of the respective geometry showing good accordance with the proposed effects. A simplified physical model is proposed for both instruments along with a more complete physical model taking the geometry of the sound production mechanisms of the instruments into account.

F. Pfeifle (✉) · M. Münster
Institute of Systematic Musicology, University of Hamburg,
Neue Rabenstrasse 13, 20354 Hamburg, Germany
e-mail: florian.pfeifle@uni-hamburg.de

M. Münster
e-mail: m.muenster@arcor.de

A. Schneider (ed.), *Studies in Musical Acoustics and Psychoacoustics*,
Current Research in Systematic Musicology 4,
DOI 10.1007/978-3-319-47292-8_3

1 Introduction

The technological advances in the 19th century that put a mark on many areas of human culture and modern living had, and still have, an formative influence on music production, processing and perception. The utilisation of principles from natural science for sound producing as well as sound modification purposes has a long tradition in different musical styles and genres of the 19th and 20th century. And both areas, music and science, influenced each other in several regards. There are several illustrious examples of electromechanical effects being utilised in the tone production of music instruments, see for instance Hammond organs or early synthesizers.

Today, the majority of keyboard instruments make use, more or less, of digital sound generation, either utilising special sound producing chips or using sampled sound libraries. Nonetheless, and this is somewhat remarkable, many of the digital sounds available in modern keyboards and synthesizers are based on analog instruments either completely acoustic, electro-mechanic or analog-electronic, pointing to a certain preference by musicians as well as music consumers. Thus, a faithful reproduction of those originally analog sounds can help to enhance the musical as well as artistic experience of such sound synthesis methods. The electro-mechanic effects on the other hand can be used to illustrate physical principles of such tone-production, and pickup mechanisms, showing how the characteristic timbre of such instruments is created by utilising fundamental principles of electrodynamics.

In this treatise, two idiomatic examples of electromechanical keyboard instruments are presented. Among two of the most popular “E-Pianos” are the Wurlitzer EP200 and the Rhodes Mark-I/Mark-II pianos, still highly valued among musicians, music producers and evoking specific associations among listeners regarding their specific genre, which primarily is Jazz, Funk and Soul music.

Throughout the following pages, a focus is put on the primary sound production mechanism of both instruments and it is shown that their characteristic timbre is due to the specifics of the respective conversion mechanism of the mechanical motion into an electronic signal, in both cases an alternating current. The influence of the electronic circuit following the basic sound pickup system is left out of the consideration here because the most characteristic part of the instruments sound is produced at the pickup mechanism as will be shown in the following.

The acoustic research history on both instruments is comparably sparse [1,2] and the effects which are published in patent specifications of the respective instrument omit some specific properties of the mechanism and an influence of certain parameters [3]. In this treatise we want to elucidate the mechanisms to aid the development of a physical model for sound synthesis and auralisation of both instruments.

After a short historic overview, the physical effects of both tone production mechanisms are described and a series of measurements are presented along with a consideration of the influences of the investigated effects. These are combined to a simplified model of the instrument, implemented using finite difference schemes

and a more elaborate model of both instruments taking the specific geometry of the instruments pickup mechanism into account.

2 History

In this section a short overview on the history and the evolution of both instruments is given, a focus is put on the inventions surrounding the primary sound production of the instruments.

2.1 History of the Rhodes

An early electromechanical instrument was constructed by Thaddeus Cahill (1867–1934) in 1906. The *Dynamophone* or *Telharmonium*, a vastly huge organ instrument with motor driven wheels having different profiles. The rotating, later called tonewheels induce a change in voltage in a magnetic field of a wire coil around a permanent magnet, following to their profiles. This idea delivered the conception for Laurens Hammond's successful organs. It is directly referable to the principle of the AC-generator from 1832 by Antoine-Hippolyte Pixii (1808–1835).

The first commercially successful application was an electrical phonograph pick up, introduced in the 1920s. The first obtainable musical instrument with such a pick up were Rickenbacker's Hawaiian lap guitars A22/A25 (known as Frying Pan), developed by George D. Beauchamp 1932 [4, 5]. The earliest known piano like instrument using an electromagnetic pick up was the Neo-Bechstein piano, a modified acoustic grand piano using pickups to capture string motion and subject it to electronic modification and amplification. It was conceived by Walther Nernst in 1930, together with the companies Bechstein and Siemens. In 1940 Earl Hines started touring with a RCA Storytone Electric Piano, a comparable construction being sold in the U.S. [6].

The Rhodes electric piano was invented by Harold Burroughs Rhodes (1910–2000). As a piano teacher he developed his own teaching method. During World War II he invented the Army Air Corps Piano to enable recovering soldiers to play piano. It was a miniaturised acoustic piano using aluminium tubing instead of strings to produce a xylophone-like sound much like a toy piano. After the war, H. Rhodes founded the Rhodes Piano Corporation to built and sell a more advanced instrument, the Pre-Piano with a new electromagnetic tone production [7, 8]. Leo Fender, already a big name in making and marketing electric guitars and amplifiers, acquired the Rhodes Piano Corporation in 1959. The first model was the Rhodes Piano Bass. The generator part now included a so called Rhodes Tuning Fork [9]. The assembly was refined to produce an intense fundamental tone, lacking higher harmonics. Under the leadership of CBS who bought out Leo Fender 1965, sales were enforced. Gaining popularity in several genres of popular music originating from jazz music,

the Rhodes piano became the largest-selling electronic piano until the end of production 1983 due to upcoming, affordable polyphonic synthesizers and samplers. Since the 1990s, the instrument enjoyed a resurgence in popularity. From 2007 on it has been reissued by the Rhodes Music Corporation as Rhodes Mark 7.

The standard models Mark-I and Mark-II did not came with pre-amplifiers, their electronics are passive, comparable to most electric guitars. Among a few works' own solutions like the Rhodes Janus I -P.A.-system, the suitcase models with built-in amplifiers and the Fender Twin Reverb, the Roland Jazz Chorus-Line of guitar amplifiers are common amplifier choices for stage and studio.

2.2 *History of the Wurlitzer*

In contrast to the history of the Rhodes Piano the conditions and circumstances were different in the case of the Wurlitzer E-piano. While the first Rhodes Pianos were produced under the leadership of a young, small but famous Californian guitar manufacturer, which had no experience in making keyboard instruments at all, the Rudolph Wurlitzer Company Ltd. (1853–1985) started as a retailer of stringed, woodwind and brass instruments from Germany and supplied U.S. Military bands. In 1880 the company began manufacturing acoustic pianos. Later, they were very successful in making band organs, *orchestrions*, *nickelodeons*, jukeboxes as well as theatre organs.

Most tone production mechanism of the aforementioned instruments are based on mechanical principles, whereas the Wurlitzer E-piano series makes use of an electrostatic pickup system. The company had some preliminary experience in the use of this technique. After world War II, Wurlitzer acquired the Everett Piano Company who manufactured the *Orgatron* which was an electrostatic reed organ developed by organist and conductor Frederick Albert Hoschke in 1934. Wurlitzer kept the Orgatron in production until the mid-1960s. The pickup mechanism of Wurlitzers electrostatic organs and pianos lie in the same plane as the vibrating reed, opposed to the U.S. patent which includes a description of such a construction with extended "ear like" metal plates [10], whereas later models omit these "ears". The principles of electrostatic pickups were patented by Benjamin F. Miessner in the U.S. [11] and at about the same time by Oskar Vierling in Austria and Germany 1932 [12, 13]. Their supposedly common research led to the *Elektrochord*, a string based piano with electrostatic pickups [6].[1]

[1]There are speculations about earlier electrostatic pickup system supposedly developed by sound engineer and luthier Lloyd Loar while he worked for the Gibson Guitar Company from 1919–1924. There are no designs preserved from this time nor are there schematic drawings which would substantiate this assumption [14] but at least one of the original Loar-designed L5s from 1929 was factory fitted with an electrostatic pickup [15] which was incorporated into the resonance body of the guitar thus picking-up only body vibrations and not the vibrations of the strings as electromagnetic pickups do.

The first electronic piano consisting of a comparable pickup system as modern Wurlitzer E-pianos was the Model 100 marketed in 1954 [16]. This early instrument was followed by a series of similar models, the EP110, EP111, EP112 which had several small differences and enhancements compared to the earliest model but had a similar tone production mechanism. All Instruments had pre-amplifiers and small power amplifiers to drive built-in speakers. The pre-amplifier includes patented high- and low-pass filtering; later transistorised models had local negative feedback within the circuitry to suppress system immanent noise produced by the sensitive tone generator. Including all variations and different sub-models there supposedly exist between 40 and 50 models differing in shape, size and/or amplification circuitry [17]. Among these, the most popular Wurlitzer model, manufactured until 1981, is the Wurlitzer EP-200A. This model, which typically consists of a black plastic body, incorporates an amplifier and two small speakers built into the casing and facing the player. Similar to earlier models it consists of a tremolo sound effect which can be gradually added to the amplified sound of the instrument.

3 Physical Properties

In this section an overview of the physical properties of the instruments measured in this work is given. A focus is put on primary sound production mechanisms of the Rhodes and Wurlitzer electronic pianos and their respective tone production geometries.

3.1 Sound Production of the Fender Rhodes Electric Piano

The sound production of the Fender Rhodes piano can be divided into two parts, a mechanical part and an electromagnetic part.

The mechanical part consists of a rod made of spring steel shrunk into an aluminium block on one side, making the resulting system comparable to a cantilever beam. The length and circumference of the rod as well as the position of a small tuning spring, adding mass, determines its fundamental frequency. The rod, which in the case of the Rhodes piano is called a tine, is excited by hammer that has a neoprene tip. The key action mechanism is a simplified single action as described in [18], it can be compared to a *Viennese* or *German* piano action because the hammer is in direct contact with the key. Depending on the year of construction the key and hammer mechanisms are crafted from wood or, as generally used in newer models, of synthetic materials. Every tine is damped by an individual felt damper that is in contact with the tines from below. The fixation of the tine, the aluminium block, is tightly connected to a, sometimes $\frac{pi}{2}$ twisted, brass bar which acts as the second prong of the patented Rhodes' "tuning fork" system.

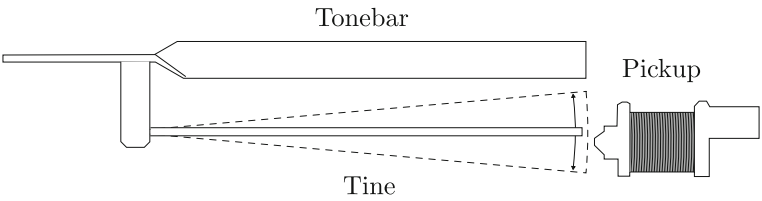


Fig. 1 The Rhodes tuning fork assembly with electromagnetic pickup

When played softly, the sound of a Rhodes piano can be described as *glockenspiel*-like, consisting of an extremely short transient showing higher, non-harmonic partials quickly leading to a quasi stationary waveform after 10–30 ms. As shown in (isma/jasa), the higher non-harmonic partials are created by the brass bar and are more prominent in the upper register of the instrument.

Depending on the velocity of the attack, lower notes tend to have a rich harmonic sound characteristic, often described as a "growling" sound. By playing more softly, the fundamental frequency of the tine vibration is more present, by gradually increasing the playing force the sound becomes successively more "growly". This playing characteristics adds to the Rhodes piano's expressivity as a music instrument.

The harmonic oscillations of the mechanic part of the Rhodes' tone production is converted to an alternating voltage by an electromagnetic pickup, that consists of a wound permanent magnet comparable to a pickup of a guitar in its overall structure but differing in terms of the magnets geometry as is depicted in Fig. 1. It consists of a round ferrite permanent magnet attached to a frustum made of iron. The magnet is wound by a coil consisting of ≈2500–3000 turns of 37 AWG[2] enamelled wire running on a synthetic bobbin.

The geometry of the pickup's iron tip shapes the specific distribution of the magnetic field in which the tine vibrates. The motion of the ferromagnetic tine changes the flux of the magnetic field which in turn produces a change in the electromotive force of the pickup resulting in an alternating voltage which then can be amplified by an external amplifier. The copper wire winding of each pick up is divided into two sections, connected in opposite phase for hum cancelling. The sound of a tone can be altered by changing the position of the tine in respect to the magnet. The more a tine is aligned towards the center of the wedge shaped magnet the more symmetrical the resulting waveform is. When aligned perfectly centered, the produced sound behind the pickup is twice the fundamental of the tine as schematically depicted in Fig. 17a. The more the tine is shifted towards the edge the more asymmetric the resulting sound is, leading to a higher amount of harmonic partials which is classified as "growl" by most musicians. The influence of this effect is represented in Fig. 17b. In higher registers, the Rhodes' tine is smaller thus having a smaller deflection which results in a smaller change of the magnetic flux

[2]American Wound Gauge.

and the resulting sound has a stronger fundamental without a comparable amount of higher partial as lower notes tend to have.

3.1.1 Measured Instrument

The instrument measured for this treatise is a Rhodes Mark-II stage piano consisting of 73 keys. It is equipped with synthetic hammers with a neoprene tip which is the typical material choice for Rhodes E-piano hammers since the mid-70s. The keys themselves are made of wood.

3.2 Sound Production of the Wurlitzer EP300

In contrast to the Rhodes' electromagnetic pickup system, the Wurlitzer piano sound production utilises electrostatic effects. A steel plate that is impacted by a hammer vibrates as an electrode of a capacitor resulting in a time varying capacitance. The plate, called reed in the user manual of Wurlitzer pianos [19], is made of hardened light spring steel, fixed at one end and free at the other. There are two factors determining the fundamental frequency f_0 of every reed, the physical dimensions of the reed itself and the amount of solder on the tip of the reed. By removing or adding lead to the tip of the reed its f_0 is increased or lowered respectively. As shown in Fig. 2, a voltage of 170 V is applied to a stationary plate and the reed acts as the low potential electrode of the resulting capacitor. The charged plate has cutouts at the position of the reed for each note of the instrument. The reeds are able to vibrate freely between the symmetric cutouts, providing a surface area large enough to produce a measurable change in capacity. The air gaps between plate and reed act as dielectric material. Analogous to the case of a plate capacitor or the diaphragm of a condenser microphone, the capacity varies inversely proportional to the distance between the two electrodes, here, reed and fixed plate.[3]

The key action mechanism of the Wurlitzer piano consists of a miniaturized London style piano action that can be regulated like a grand piano action. Every reed of the Wurlitzer piano is excited by an individual ply maple hammer that has a felt tip [19]. Comparable to the playing dynamics of the Rhodes E-piano, depressing the keys with higher velocity results in a richer harmonic sound of the Wurlitzer than playing softly.

[3]As side note it should be mentioned that Miessner proposed an electrostatic pickup using high-frequency AC to pre-load the capacitor system to avoid non-linear distortion of large displacements of lower sounding, larger reeds, Wurlitzer instead choose to set the DC pre-load high enough to keep the E-field large to prevent effects of distortion.

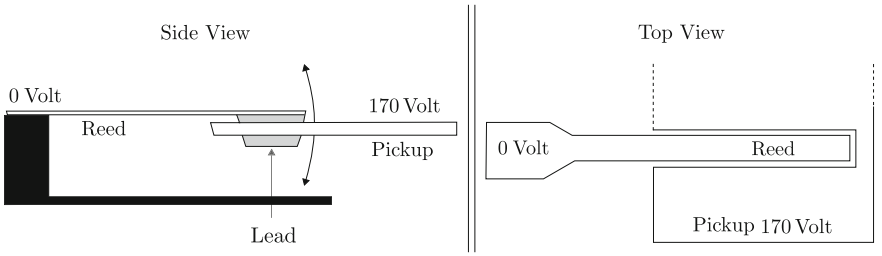


Fig. 2 Structural depiction of the Wurlitzers pickup system. A side view on the *left*, top view on the *right*. Both showing the high potential plate and the low potential reed

3.2.1 Measured Instrument

The Wurlitzer piano model measured in this treatise is a Wurlitzer EP300 which has an amplification circuitry which is a mixture of the popular EP200A and the EP140 and is a model which was only marketed in Germany. The tone production is similar to the EP200 series, but, in comparison to the synthetic case of the EP200A it consists of a ply wooden case containing three integrated speakers as well as individual inputs and outputs for head-phones, external speakers or microphones. Comparable to most Wurlitzer piano models it consists of 64 keys ranging from A1 with a fundamental frequency of 55 Hz to C7 with a fundamental frequency of 2093 Hz.

Contrary to the values given in Wurlitzer's service manual schematics, a measurement of the high potential plate of this instrument shows ≈147 V and not 170 V, as indicated in the manual. Shown in Fig. 3 is the resistor where the direct-out voltage is measured. The physical properties of the reeds of this instrument are given in Fig. 4 and Table 1.

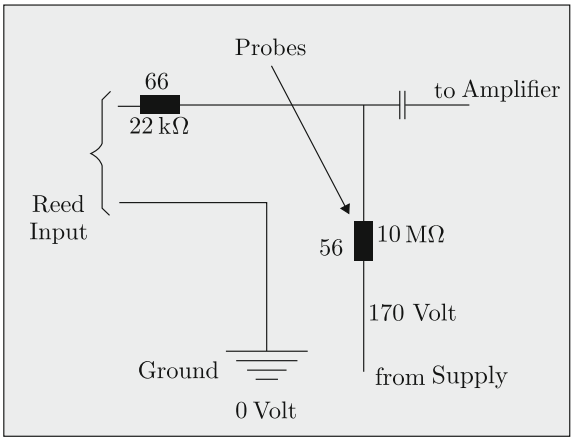


Fig. 3 Section from the Wurlitzer EP300 schematic. Indicated by the *arrow* the resistor where the electric probe measurements are performed

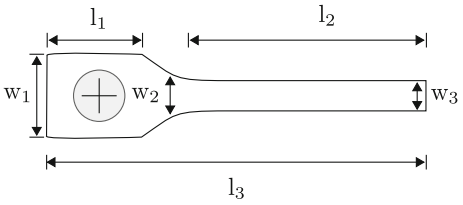


Fig. 4 Length indexes for the values given in Table 1. The speaking length l_2 is the portion of the reed not in contact with its mounting

Table 1 Physical sizes of Wurlitzer reeds

	Base length (l_1)[a]	Speaking length (l_2)	Total length (l_3)	Total width (w_1)	Base width (w_2)	Tip width (w_3)
Lowest reed	13	61	74	10	6.7	4
Highest reed	13	13	27	10	2.8	2.5
measured reed	13	47	62	10	6.5	3.8

[a]All values in [mm]

The tines are fixed by screws at their base shown in Fig. 4. The contact point between hammer and the Wurlitzer's reeds is approximately at half the reed's speaking length. The impact point between hammer and the measured reed is at 22.27 [mm] from the tip which is approximately half its speaking length.

4 Methods

To characterise the exact influences of different parts belonging to the tone production, a set of measurements are performed using different methods. All measurements presented here are carried out at the Institute of Systematic Musicology at the University of Hamburg.

4.1 Camera Tracking

A high-speed camera is used to qualitatively record visibly moving parts of the instrument, and to track specific motions of the respective parts. In the case of the Rhodes E-piano, the motion of a freely vibrating tine as well as a hammer impacted tine vibration. In the case of the Wurlitzer EP200, the motion of a hammer impacted reed vibration is recorded and tracked. For all measurements, a *Vision Research Phantom V711* high-speed camera is applied. For recording and qualitative evaluation of the high-speed recordings, the Vision Research *Phantom Camera Control*

Fig. 5 A typical section of a high speed camera recording setup showing part of the Rhodes' tine including the tuning spring and the electromagnetic pickup. The tip of the tine is marked with *black ink* to facilitate motion tracking. In a realistic scenario, gain, luminosity and contrast of the camera recording are changed to emphasize tracked points

software version 1.6 and 2.7 is used. For evaluating the recording quantitatively, the *Innovision Systems* motion tracking software *MaxTraq2D* is applied. The traced trajectories are exported to an ASCII-format file, and analysed with scripts coded in *julia* language, using wavelet methods as well as Fourier transform analysis.

The measurement setup consists of the high-speed camera, a set of LED lamps and the device under test which is marked with white or black ink at several points on the geometry to facilitate automatic motion tracking.

A typical image section from a measurement is depicted in Fig. 5.

4.2 Audio Measurements

Audio signals are measured directly after the primary tone production mechanism. The Rhodes piano is equipped with a direct out jack behind the magnetic pick ups. This jack is connected to an audio recording system on a personal computer, recording the alternating voltage with a sampling rate of 44,100 Hz and 24 bit resolution.

The Wurlitzer EP200 does not consist of an output in front of the amplifying circuit, thus, the voltage is measured over a resistor using an electric probe which is connected to a high-precision measuring amplifier and converter sampling at a frequency of 50.0 kHz with a bit depth of 24 bits. The specific resistor is indicated in Fig. 3.

5 Measurements

The measurements are performed on a Fender Rhodes Mark-II and a Wurlitzer EP300. The vibrational behaviour of the sound production assemblies are investigated using high speed camera techniques and audio recordings of the instrument sound immediately following the electromechanical pickup system.

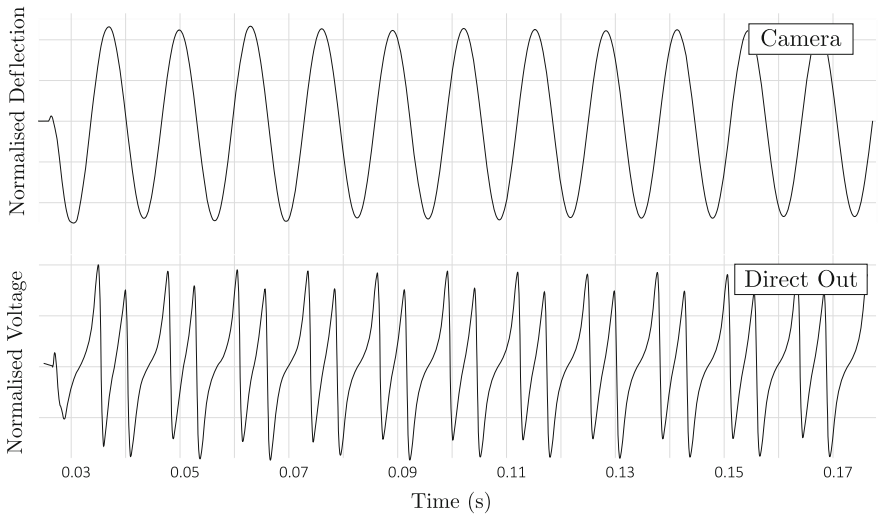


Fig. 6 The *upper graph* shows the tracked signal from a high-speed camera recording of the tine's tip, exhibiting approximately sinusoidal motion. The *lower graph* shows the voltage measured behind the pickup at the direct-out jack of the Rhodes Stages piano

5.1 Rhodes

To characterise the influence of the Rhodes' tine on the resulting sound, several tines are measured using high-speed camera recordings. Figure 6 shows the tracked motion of a Rhodes tine tip with the fundamental frequency of ≈78 Hz and the resulting direct-out sound recorded behind the pickup of the same tone. The note is played *forte*.

The measured signals show that the primary vibrating part of the Rhodes' tone production, the tine, is vibrating in almost perfect sinusoidal motion. The direct-out measurement shows a considerably more complex behaviour pointing to the fact that the magnetic pick up is the main contributory factor of the specific instrument sound. As depicted in Fig. 7 the spectrum of the measured audio signal shows rich harmonic content with a smooth decay of the higher partials and a long-lasting fundamental. A small amount of beating is visible in the first harmonic around the 3 s mark and also in the 4th and 6th harmonic.

To classify the influence of the hammer impact four points in the vicinity of the contact area between hammer tip and tine are recorded and tracked. Figure 8 shows that the hammer impact lasts approximately 4.7 ms and is divided into one longer period and a short reflection, this behaviour is comparable to the occurrence of multiple contacts in low register piano string/hammer excitation []. Comparable to the vibration characteristics of the tine tip, a measurement near the impact point shows sinusoidal motion after approximately one cycle.

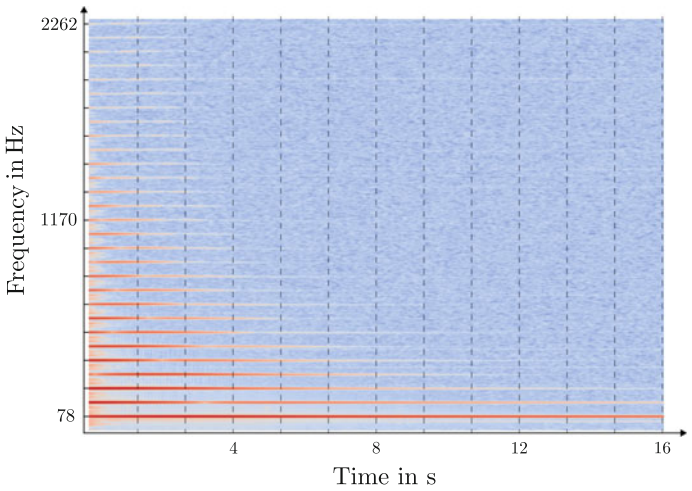


Fig. 7 A spectrogram of the measured audio signal of the Rhodes

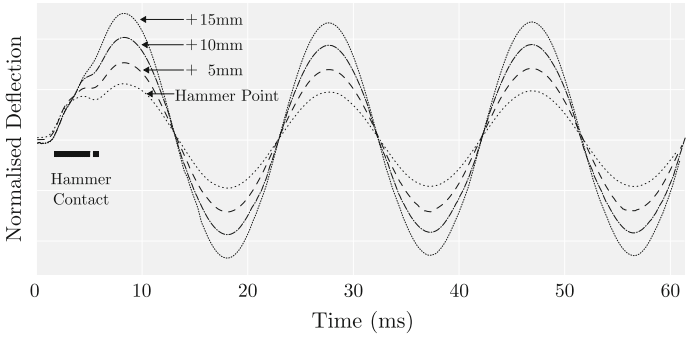


Fig. 8 Four tracked points near the impact zone of the hammer. The *black bar* indicates the contact time between hammer and tine

5.2 Wurlitzer

In this measurement setup, the tip of a Wurlitzer reed excited by a forte stroke is recorded and tracked. Figure 9 shows the tracked motion of a Wurlitzer reed tip with a fundamental frequency of ≈98 Hz under normal playing conditions and the resulting direct-out sound of the same measurement. Corresponding to the measurements of the Rhodes piano, the tip of the Wurlitzer's reed shows an approximate sinusoidal motion whereas the sound recorded behind the pickup exhibits a considerably complex wave form. Again pointing to the fact that the electrostatic pickup is essential for the formation of the specific Wurlitzer sound. As shown in Fig. 10 the recorded audio signal shows a highly complex spectrum with up to 40

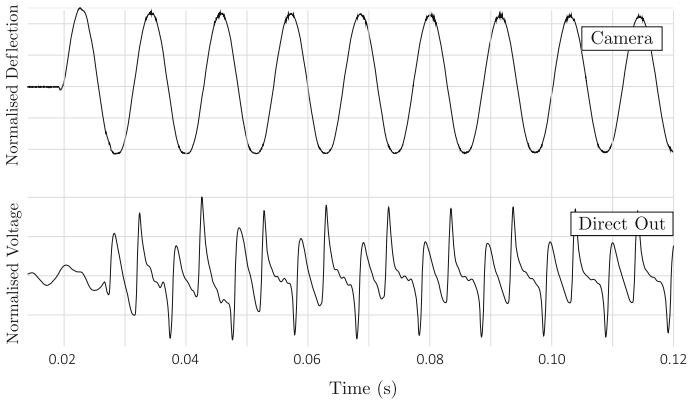


Fig. 9 The *upper graph* shows the tracked signal from the high-speed camera recording again exhibiting approximately sinusoidal motion. The *lower graph* shows the voltage measured behind the pickup over a resistor ahead of the pre-amplification circuitry

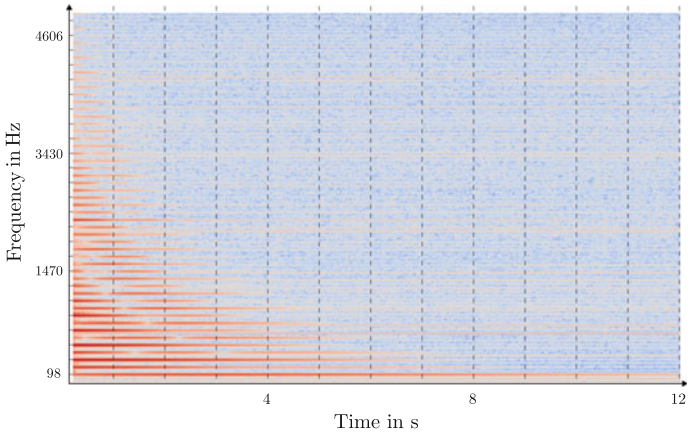


Fig. 10 A spectrogram of the measured audio signal of the Wurlitzer

partials present in the first second of the sound. In addition to the rich harmonic content, the decay characteristics of higher partial show a complex non-exponential decay behaviour with several partials showing a strong beating, e.g. the 3rd and the 5th.

The influence of the hammer is tracked at several points around the impact position. Figure 11 shows that the hammer has a small but noticeable influence on the measured vibration. And the motion is not immediately sinusoidal like the Rhodes tine. The hammer impact lasts around 1.25 ms.

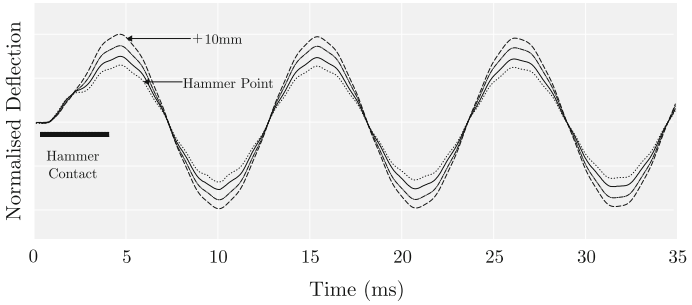


Fig. 11 Four points near the impact zone of the hammer. The *black bar* indicates the contact time between hammer and tine

6 Intermediate Results

The presented measurements of the mechanic part as well as the electronic part of the tone production of both instruments leads us to the intermediate conclusion that the primary mechanical exciters play only a secondary role in the sound production of both instrument and the specific timbre is influenced more by the specific pickup system. In particular the electromagnetic pickup of the Rhodes and the electrostatic pickup of the Wurlitzer. A crucial part of the instruments sound characteristic and timbre must thus be attributed to the coupled electro/mechanical systems at hand. The measurements of both instruments show that the main vibrating parts are vibrating approximately in sinusoidal motion. The resulting sounds measured directly behind the electrostatic or electromagnetic pickup show a more complex behaviour. In the case of the Wurlitzer, the specific pickup geometry leads to a highly complex decay characteristic showing interesting effects like non-exponential decay characteristics and beating of higher partials.

7 Finite Element Models of Sound Production Assemblies

To assess the influence and the specific distribution of the magnetic and electrostatic fields in the vicinity of the pickups [20–22], finite element method (FEM) [23] models of the sound production units of both electric pianos are developed and simulated using the FEM tool and solver *Comsol Multiphysics*.

7.1 Magnetic Field of the Rhodes Pickup

The FEM-model of the Rhodes' pickup system includes the magnetic field surrounding the iron conic section as well as the attached magnet. It is simplified by

omitting the copper coil windings and thus leaving electrodynamic effects out of the consideration. The static magnetic field distribution is computed using a scalar magnetic potential. Assuming a current free region, the relation

$$\nabla H = 0 \tag{1}$$

holds, with H being the magnetic field. Comparable to the definition of the electric potential for static E-fields, the magnetic scalar potential V_m is given by

$$H = -\nabla V_m \tag{2}$$

Using the equivalence $B = \mu 0(H + M)$, and $\nabla B = 0$, where B is the magnetic flux density we can rewrite Eq. 2 to

$$-\nabla(\mu_o \nabla V_m - \mu_o M) = 0\ , \tag{3}$$

with M the magnetization vector describing the magnetization of a material influenced by magnetic field H. Generally M can be seen as function of H [24, pp. 195 ff].

The tine of the Rhodes is positioned in close proximity to the steel tip of the pickup. The flattened sides of the frustum focuses the magnet field in the center showing an approximate bell curve characteristic. The sound is shaped by the distance between the tine and the magnet, caused by the strength of magnetic flux at the respective position. The model shows the disturbance of the magnet field [25]. As the deflection of the tine gets larger, it leaves the magnet field resulting in a more asymmetrical change magnetic of magnetic flux. An idealised model of the pickup system is depicted in Fig. 12 showing a distribution of the static H-field forces surrounding the tip of the magnet.

Geometry of the Pickup Tip

To classify the influence of the pickup shape three models with different tip geometries are created. The resulting H-field in the normal direction of the pickup on a curved line approximately 8 mm above the tip are depicted in Figs. 13 a–c.

As is depicted in Fig. 13a–c, the specific form of the Rhodes' pickup shapes the magnetic field in front of the pickup resulting in a bell shaped curve with different Q-factors.[4]

7.2 *Electrodynamic Interaction of the Wurlitzer Piano*

The FEM model of the Wurlitzer pickup system is developed to solve the dynamic influence of the vibrating reed on the capacitance of the quasi-condenser system. This is achieved by solving Poisson's equation for several static positions on the

[4]The Q-factor is defined as the ratio of the center frequency and the bandwidth. In our case, the center frequency is the position above the symmetry axis of the magnet's tip.

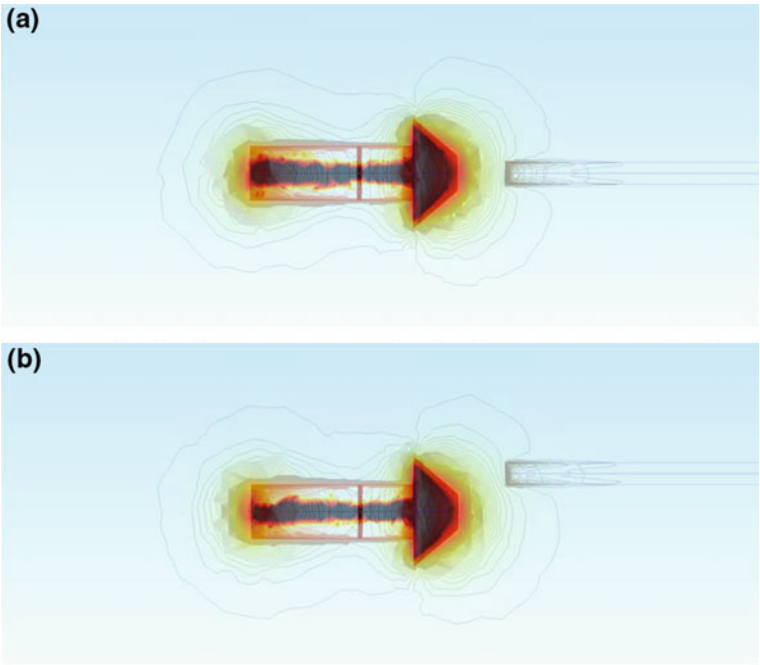


Fig. 12 A FEM simulation of the Rhodes' tine and pickup system showing the resulting force lines due to the magnetic field. The tip of the tine is magnetised as well which is indicated by the force lines on the tine. **a** Symmetric positioning in front of the magnet. **b** Asymmetric positioning of the tine in front of the pickup

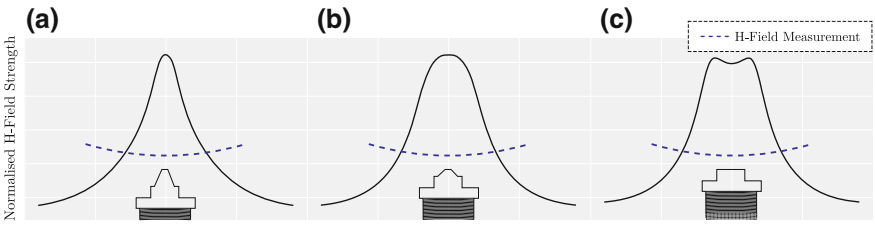


Fig. 13 A FEM simulation of the Rhodes' pickup tip and the resulting H-field strength on a *curve* above the tip. **a** A narrow and high pickup tip and the resulting H-field. **b** A medium high pickup tip and it's resulting H-field. **c** A planar pickup tip comparable to the top of a guitar pickup and the resulting H-field

trajectory of the reed's motion. The stationary electrode of the modeled pickup is charged with a voltage of 147 V whereas the reed is kept at zero potential. The pickup behaves similar to a plate capacitor, where changing distances over time between the reed and the plate results in a changing capacity. A post-processing

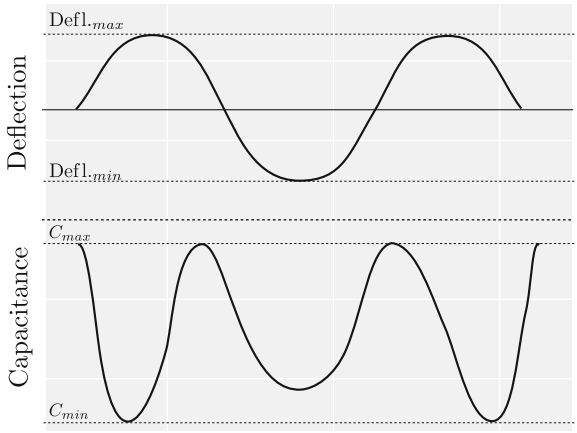


Fig. 14 The changing capacitance due to changing deflection of the Wurlitzer reed

step then computes the change in capacitance. The models work under the following assumptions:

The electric scalar potential, V, satisfies Poissons equation:

$$-\nabla(\varepsilon_0 \varepsilon_r \nabla V) = \rho \tag{4}$$

ε_0 is the permittivity of free space, ε_r is the relative permittivity, and ρ is the space charge density. The gradient of V gives electric field and the displacement:

$$E = -\nabla V D = \varepsilon_0 \varepsilon_r E \tag{5}$$

Boundary conditions of the electric potential are applied to the reed and plates. A potential of 147 V is applied to the plate, whereas the reed maintains grounded. For the surrounding air of, conditions corresponding to zero charge are applied:

$$n \times D = 0 \tag{6}$$

The capacitance of the pickup is changed by the motion of the moving reed. A varying current flows into the plate as needed to maintain the proper charge for a new amount of capacitance. This current produces a varying voltage across an external resistor which is decoupled and amplified to produce a usable output signal as shown in Fig. 3.

The changing capacitance is depicted in Fig. 14. At the capacitance minima of the curve, the excitation of the reed is maximum and at the peaks where capacitance is maximum the reed is near its rest position. Because of the non-symmetric design of the Wurlitzer's reed, the capacity change differs at each excursion depending on moving direction as already measured in 1965 by Ippolito at Wurlitzer Co. [26].

8 Finite Difference Models

The numerical models presented in this section are based on the measured properties presented in Sect. 5, qualitative observations of the FEM models presented before and some conjectures regarding material properties of the Wurltizer's reed and the hammer tip of both instruments. Taking the measurement results as a basis for the models results in several assumptions that simplify the model description of the physical system considerably. Regardless of the introduced simplifications both models are able to capture the vibratory motion and the acoustic properties of both instruments to a high degree while minimizing computational as well as modeling complexity.

A model of both pickup systems including all physical parameters would have to take time-varying electromagnetic effects into account using Maxwell's equations for electromagnetism to describe the respective pickup mechanism in complete form. But, due to the small changes in the magnetic as well as electric fields the proposed simplifications lead to models that are able to approximate the vibratory as well as the sonic characteristics of the instruments very accurately.

The Rhodes models presented here is an extension of the model published in (ISMA 2014) and corrects several shortcomings and imprecise assumptions of this earlier work. The model of the Wurlitzer EP200 shares conceptual similarities with the Rhodes model but is adapted to the different geometry of the sound production. Both models consist of a hammer-impacted resonator exiting a spatial transfer function modeled after the characteristic pickup system of the respective instrument.

8.1 Rhodes Exciter Model

As shown in Fig. 1 the tip of the tine vibrates in close proximity to the electromagnetic pickup and the FEM simulations given in Fig. 12 show that only a small part of the tip is influenced by the magnetic field. Therefore, the exciter of the Rhodes is modeled as a hammer impacted simple harmonic oscillator (SHO) representing the quasi-sinusoidal motion of the tip.

Using Newton's second law, the temporal evolution of a SHO can be written as a second order ordinary differential equation

$$x_{tt} = -\kappa \cdot x \tag{7}$$

with $\kappa = \frac{k}{m}$ the stiffness/springiness of the system, m the mass of the harmonic oscillator, x the deflection and the subscript by t on the left hand side indicating a second derivative in respect to time.

A hammer impact with elastic material properties of the hammer tip can be simulated by using a hysteretic hammer model as presented in [27, 28]. This impact

model is able to simulate hammer impacts of different materials showing viscoelastic behaviour.

Thus, the impacted SHO is extended to

$$x_{tt} = -\kappa \cdot x - F_{\text{int}} \tag{8}$$

with F_{int} the resulting contact force between hammer and SHO, following [27], this force follows the relationship

$$F_{\text{int}}[x(t)] = \begin{cases} k \cdot x(t)^{\alpha} + \lambda \cdot x(t)^{\alpha} \cdot x_t(t) & \text{if} \quad x > 0 \\ 0 & \text{for} \quad x \leq 0 \end{cases} \tag{9}$$

which originally is a model for hammer impacts developed by Hunt and Crossly [29], that has shown to yield good results for models of hammer impacts with moderate impact velocities and plain geometries [27, 30]. Here, α is the nonlinearity exponent depending on the geometry of the contact area and λ is a material dependent damping term that dissipates energy in dependence to the velocity of the changing hammer-tip compression written as x_t.

A typical hammer force over hammer-tip compression curve is plotted in Fig. 15.

The differential equation for both systems can be separated by defining $v = u_t$, the velocity and thus rewritten as

$$\begin{aligned} v_t &= -\kappa \cdot x \pm F_{\text{int}} \\ x_t &= v \end{aligned} \tag{10}$$

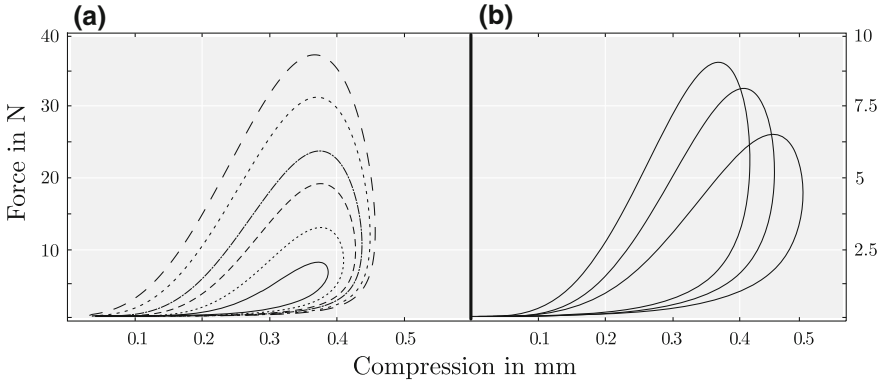


Fig. 15 Force over compression profiles for different hammer parameters. **a** Different values for damping constant λ. **b** Different values for for non-linearity exponent α

8.1.1 Finite Difference Approximation

The exciter models of the Rhodes and the Wurlitzer pianos are discretised applying standard finite difference approximations using a symplectic Euler scheme for iteration in time. The discretisation method and the scheme are published in more detail in [31, 32]. Applying standard FD approximations for the given problem using the operator notation given in Appendix II and iterating the scheme in time by using mentioned method leads to two coupled equations

$$\begin{aligned} \delta_t v_{sho} &= -\kappa_{sho} \cdot x_{sho} - \gamma \delta_t x_{sho} - F_{int} \\ \delta_t x_{sho} &= v_{sho} \end{aligned} \tag{11}$$

for the impacted SHO and

$$\begin{aligned} \delta_t v_{ham} &= -\kappa_{ham} \cdot x + F_{int} \\ \delta_t x_{ham} &= v_{ham} \end{aligned} \tag{12}$$

for the hammer, with $\kappa_* = \frac{k}{m}$ the stiffness to mass quotient of the SHO and the hammer respectively. Equation 11 consists of a viscous damping term which heuristically approximated damping parameter γ. The interaction force is computed by relation 17.

8.2 *Wurlitzer Exciter Model*

The reed of the Wurlitzer is modeled as a cantilever beam including large deflection effects, modeled by the inclusion of shearing effects in the beam. Trail and Nash [33] showed that the shear beam is a better approximation for the vibrations of the fundamental frequency then the more popular Euler-Bernoulli beam and less computationally complex than the similar accurate Timoshenko beam model.

The use of a beam model instead of a plate model is justifiable here because torsional motion of the plate were not measured using the high-speed camera setup and thus are either not present or small compared to the transversal deflection of the fundamental mode. In addition to that, the measurements show that the influence of higher modes are comparably small and the mode of vibration could by approximated by the reeds first natural frequency.

The decision to model the vibration of the reed as a 1-dimensional geometry is due to the fact that a larger part of reed influences the electrostatic field as visible in Figs. 2 and 18.

Compared to its height, the deflection of the Wurlitzer's reed is large. Thus it is feasible to include high deflection effects into the formulation of the model. As shown in [34] the inclusion of shear effects to the Euler-Bernoulli beam raises the accuracy of the fundamental frequency as well as the accuracy of higher partials.

Without further derivation we introduce the formulation of the shear beam as developed in [34]. By separating both dependent variables, the deflection and the angle, the equations of transversal motion for a shear beam, given as a partial differential equation can be written as

$$\frac{1}{\rho \mathrm{A}} u_{tt} - \frac{1}{k' \mathrm{GA}} \cdot u_{4x} - \kappa u_{2x2t} - f(x,t) = 0 \tag{13}$$

with $\rho, \mathrm{A}, \mathrm{G}$ dimensionless variables given in Appendix I. Equation 13 does not explicitly depend on the shear angle α (see [34]) thus it is not regarded here any further. Again omitting the shear angle, the boundary conditions for the fixed/free beam are

$$\begin{aligned} u|_0 &= 0 \\ \mathrm{k}' \mathrm{GA} u_x|_L &= 0\ . \end{aligned} \tag{14}$$

8.2.1 Finite Difference Approximation

Again introducing $v = u_t$ and using finite difference operators as defined in Appendix II, it is possible to reduce the PDE 13 to two coupled ordinary differential equations (ODE) thus rewriting the problem as

$$\begin{aligned} \delta_t v &= [\delta_{4x} - \delta_{xx}\delta_{tt}]\mathbf{u} + F([\mathbf{x}], t) \\ \delta_t u &= v \end{aligned} \tag{15}$$

and the boundary conditions as

$$\begin{aligned} u|_0 &= 0 \\ \mathrm{k}' \mathrm{GA} \delta_x u|_L &= 0. \end{aligned} \tag{16}$$

The hammer impact is modeled by using the same impact model presented 17 now including a distributed hammer force over several points on the beam indicated by

$$F([\mathbf{x}], t) = \begin{cases} k \cdot \mathbf{x}(t)^{\alpha} + \lambda \cdot \mathbf{x}(t)^{\alpha} \cdot x_t(t) & \text{if} \quad \sum_{xL} \mathbf{x} > 0 \\ 0 & \text{for} \quad \sum_{xL} \mathbf{x} \leq 0 \end{cases} \tag{17}$$

with $\sum_{xL}$ indicating a weighted sum over the contact area. The time iteration of the hammer motion is again computed by Eq. 12.

8.3 Rhodes Pickup Model

The electromagnetic effects of the Rhodes' pickup system can be reduced from Maxwell's equations for transient electromagnetic effects to a more tractable formulation know as Faraday's law of induction. As shown above, the pickup consists of a magnetized steel tip and a coil wrapped permanent magnet; leaving reciprocal magnetic effects of the induced current in the coil out of our consideration, the voltage induced over the pickup is equivalent to the change of the magnetic flux in the field produced by the magnet

$$\varepsilon = -\frac{\partial \Psi_{\mathbf{B}}}{\partial t} \tag{18}$$

with ε the electromotive force and $\Psi_{\mathbf{B}}$ the magnetic flux due to the change in the magnetic field given by

$$\Psi_{\mathbf{B}} = \int \mathbf{B} \cdot d\mathbf{S} \tag{19}$$

with B the magnetic field strength integrated over surface S. Using these equalities, the induced voltage directly depends on the change of magnetic field strength which depends solely on the position of the tine disturbing the field as shown in Fig. 12.

The following derivation of the magnetic field distribution uses the unphysical assumption that there exist magnetic monopoles which produce a distributed magnetic field.[5] As is shown in [35] this approach yields good approximations of notional magnetic induction fields produced by guitar pickups. Consisting of a plainer geometry, the tip of a guitar pickup bar magnet can be simplified to a circular, magnetically charged disc with a certain cross-section, which reduces the problem to a position-dependent integration of the field over the pickup. Due to the specific pickup geometry of the Rhodes, a different approach is taken here to calculate the induction field strength above the tip of the magnet.

As depicted in Fig. 16 our derivation makes use of several simplifying assumptions facilitating the computation.

Definition 1 The tine vibrates in an approximate sinusoidal motion in one horizontal plane in front of the pickup.

Definition 2 The tip of the tine vibrates on the trajectory of an ideal circle with the center at its fixation point.

[5]This assumption proposes an equivalence between the efficient causes of electric fields and magnetic fields and can be used as a mathematical modeling tool, see: [24, pp. 174 ff].

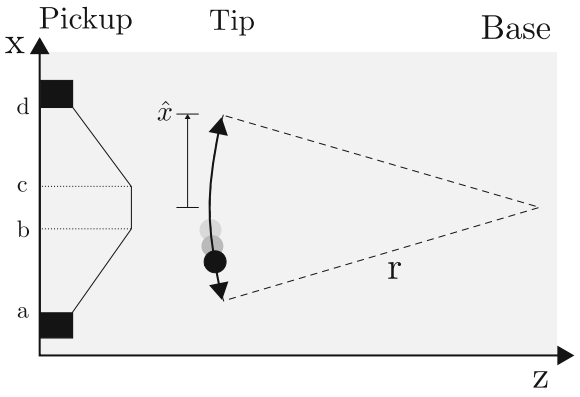


Fig. 16 Simplified geometry of the pickup system and the vibrating tine

Using Definitions 1 and 2, the calculation of the magnetic induction field depending on the position of the tine tip can be formulated as an integral over the simplified iron tip geometry.

Comparable to an electric point charge we define a magnetic point charge which produces a magnetic field given by

$$\mathbf{B} = B_0 \frac{r_{21}}{|r_{21}|^3} \tag{20}$$

with r_{21} the relative positions of the point charge and a test charge in the surrounding field. Because the magnetic flux changes only due to changes in the z direction we can reduce Eq. 20 to

$$\mathbf{B}_z = B_0 \frac{\Delta z}{|r_{21}|^3} \tag{21}$$

The magnetic field for position $(x^{'}, z^{'})$ in front of the of steel tip can thus be written as a three-part integral

$$\begin{aligned} \mathbf{B}_z(x^{'}, z^{'}) = |\mathbf{B}_{tine}| \cdot \Bigg[& \int_a^b \frac{\sigma(z^{'} - z(x))x}{[(x^{'} - x)^2 + (z^{'} - z(x))^2]^{3/2}} dx \\ & + \int_b^c \frac{\sigma(z^{'} - z_k)x}{[(x^{'} - x)^2 + (z^{'} - z_k)^2]^{3/2}} dx \\ & + \int_c^d \frac{\sigma(z^{'} - z(x))x}{[(x^{'} - x)^2 + (z^{'} - z(x))^2]^{3/2}} dx \Bigg] \end{aligned} \tag{22}$$

with σ the constant magnetic charge density.

Integrating this formula for all points on a trajectory given by the position of the Rhodes' tine tip

$$\begin{aligned} z^{'} &= \mathrm{r} - \sqrt{\mathrm{r}^2 - (x^{'})^2} \\ x^{'} &= \hat{\mathrm{x}} \cdot \sin(2\pi f_{tine} t) \end{aligned} \tag{23}$$

with f_{tine} the fundamental frequency of the tine, leads to a magnetic potential function characterising the magnitude of relative magnetic field change.

An idealised form of the magnetic field in front of the Rhodes pickup is depicted in Fig. 17a, b, it is comparable to the measurements results published in [35].

8.4 Wurlitzer Pickup Model

The influence of the pickup system of the Wurlitzer can be characterised in a similar way. Here, the change in capacitance of a time varying capacitor induces an alternating voltage which is amplified as the instruments sound.

A time-varying capacitance induces a current i

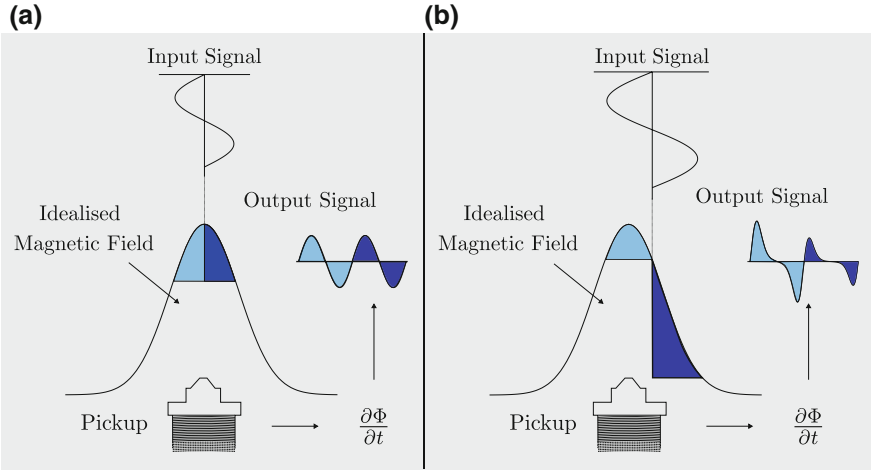


Fig. 17 An idealised schematic depiction of the pickup system of the Rhodes E-piano. The sinusoidal motion of the vibrating tine induces ac. **a** A low amplitude input of a sinusoidal vibration of the magnetic flux weighted by the magnet fields distribution. By differentiating the magnetic flux in respect to time, the alternating voltage present at the output is calculated. **b** A similar model setup as before consisting of a slightly displaced mid-point for the input motion resulting in a different weighting function of the magnetic field. The output shows a different form than before. This condition is close to a realistic playing condition found in Rhodes E-pianos

$$i(t) = \mathrm{C(t)}\frac{\partial \mathrm{u}(t)}{\partial t} + \mathrm{u(t)}\frac{\partial \mathrm{C}(t)}{\partial t} \tag{24}$$

with u the voltage and C the capacitance both depending on time t. For the derivation of the influence function of the capacitor we take two simplifying assumptions.

Definition 3 The time dependent charging/discharging curve of the capacitor is linear in the considered range.

Definition 4 The supply voltage stays constant during a capacity change cycle of the capacitor.

Using Definitions 3 and 4, we can write the time-dependent current resulting from a changing capacitance as

$$i(t) = \mathrm{u}_0 \frac{\partial \mathrm{C}(t)}{\partial t} \tag{25}$$

This alternating current induces an alternating voltage over resistor R_{56}.

To calculate the capacitance curve due to the deflection of the Wurlitzer's reed, a number of i planes through the geometry are taken and the electric field strength is computed for each resulting slice simplifying the 3-dimensional problem to a 1-dimensional. The capacitance for each slice can be computed from the electric field by

$$C_i = \frac{Q_i}{U_i} \tag{26}$$

with $Q_i = \varepsilon_t \oint_A \mathbf{E} \cdot \mathrm{d}\mathbf{A}$ the charge defined as the integral of the electric field over the surfaces of the geometries using Gauss's theorem and ε_t an electric field constant for the material and free field properties.

Three exemplary positions for the computation of the capacitance are depicted in Fig. 18.

8.5 Modeling Results

A structural flow diagram given in Fig. 19 shows that both models share similarities regarding their processing steps. Both models begin by initialising the respective values for the given geometry, then calculating the motion of the respective exciter which is then weighted with a function modeling the spatial distribution of the magnetic or electric field respectively. Both models are implemented in the high-level, high-performance language *julia* and are capable of real-time processing on a second generation medium range Intel i5 processor.

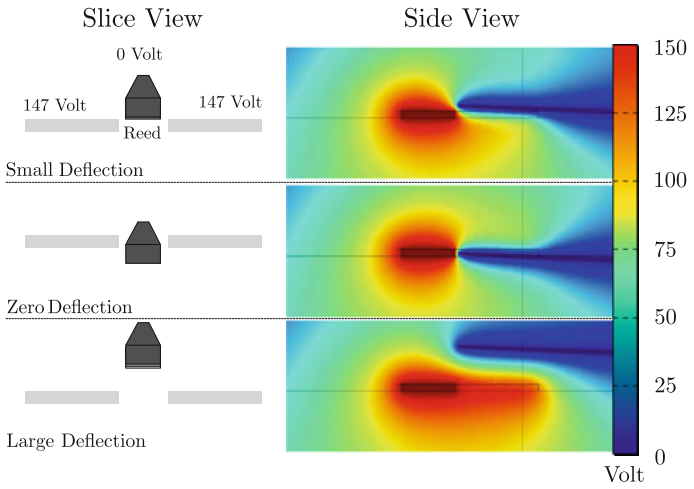


Fig. 18 Distribution of the electric field for three exemplary reed deflections. On the *left hand side* one slice of geometry on the *right hand side* the results from the FEM model

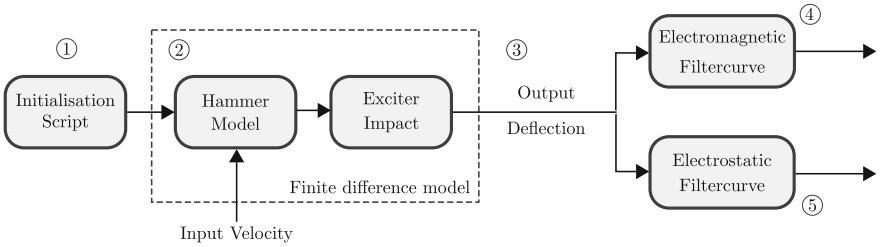


Fig. 19 Schematic depiction of the processing chain of the model. *1* The respective model is initialised regarding its physical properties and boundary condition. *2* Computation of the finite difference models. *3* Output of the respective exciter model. *4* Rhodes model output. *5* Wurlitzer model output

The simulation result for a Rhodes and a Wurlitzer E-piano tone are depicted in Fig. 20. An aural comparison of the simulated and measured sounds shows that both simulations are close to their real counterparts as can be heard on the web-site accompanying this paper.[6] In an informal listening test, which is not part of this publication, the Rhodes' sounds where rated higher than the Wurlitzer sounds pointing to the fact that the complex interaction of the Wurlitzer is approximated less well by the proposed models as the Rhodes' pickup system.

[6]More example sounds can be found on the accompaniment web-site which includes several different examples of exciter to pickup setups. See:http://www.systematicmusicology.de/?page_id=742 .

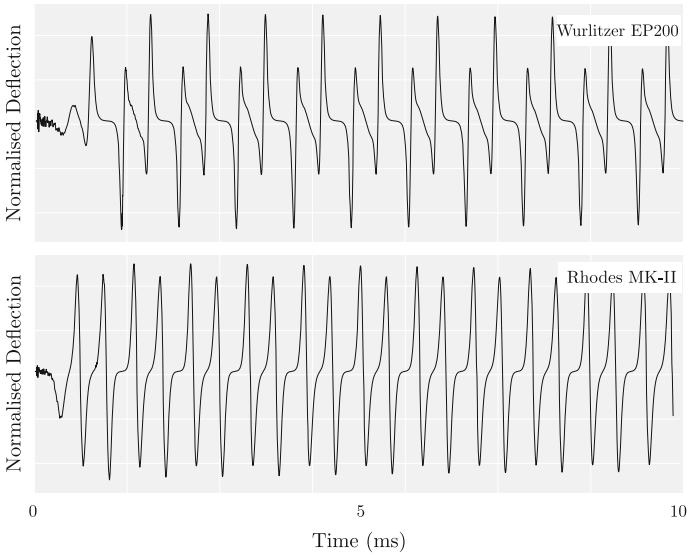


Fig. 20 The first few milliseconds of two simulated keyboard sounds. The full sounds and additional material can be found on the accompanying web-site (ref)

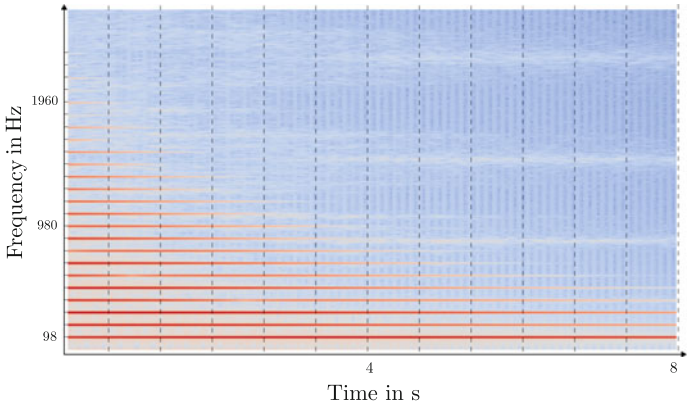


Fig. 21 A spectrogram of the simulated of the Wurlitzer sound

Inspecting the spectrogram of a simulated Wurlitzer sound given in Fig. 21 shows that there are comparably less higher harmonics in the simulated sound and the beating is not as clearly visible as in the measured sound. On the positive side, the beating of the 3rd and 5th harmonic is also present in the simulation even though it is much less pronounced.

9 Outlook

To fill the gap in acoustic research literature this paper was aimed at elucidating the basic sound production components of the Rhodes Stage Piano and the Wurlitzer EP200 series. Hence, other interesting findings are left out of this considerations and are planned to be a part of future publications. Especially interesting is the mechanism of the energy transfer between the Rhodes tine and the Rhodes bar which shows synchronisation behaviour as published in [36, 37]. At the moment, the discourse is only from a heuristic point-of-view and the development of a mathematical model for this non-linear effect is work in progress.

As already mentioned in the derivation of the pickup simulation, a complete model would call for an inclusion of time-varying effects of coupled E-fields and H-fields. Thus, a faithful simulation of these effects using Maxwell's equation of electrodynamics is a work in progress. Our hope is that the nonlinear effects of the Wurlitzer pickups can be represented with higher accuracy, and the missing effects of non-exponential decay and beating can be realised by a more complete model.

Another interesting line of research would be the mechanism of the hammer tine interaction which shows multiple contacts for low notes and high playing force. This is comparable to the effects in piano hammers known from literature.

Another route of research would be a characterisation of the influence of the electric schematics of the Wurlitzer. As the basic change of capacitance is only measurable indirectly the exact influence of the circuit is of interest for comparing the resulting sound.

Even though the Rhodes and Wurlitzer E-pianos are among the most common electro-mechanic keyboard instruments, there exist a multitude of derivatives of similar or comparable tone production principles like for instance the Yamaha CP70/CP80, Hohner *Clavinet*, Hohner *Pianet* or the Hohner *Elektra Piano* to name just a few. A comparison of the primary tone production of those other electromechanical instruments would be a fruitful topic for further considerations.

A conspicuousness that was only mentioned *en passant* in this treatise regards the question why semi-analog or analog-electronic instruments are still preferred among many musicians and listeners and are finding renewed interest over the last years. A psychoacoustic evaluation of important sound parameters in these instruments in regard to listeners' and musicians' preferences could help to answer this question.

9.1 Additional Notes on Electronics

An additional factor influencing the sound of both instruments considerably is the sort of amplification and recording techniques as several classic recordings

show us.[7] The brightness in sound we hear on these recordings are probably produced by the attached amplifiers. Electron tubes and also first/second generation transistors are known to produce a significant percentage of THD even at "clean" level settings. Nevertheless these components are designed to have nearly linear characteristic curves at least at a quiescent operation condition. Furthermore or even more important in the case of the study are aspects of power supply design. The audible compression mentioned above is encouraged by the time-current characteristic of the power supply as well [38]. Both piano sounds have a steep transient followed by a quieter decay of different length of the envelope. The current demand of the transient is to be recovered immediately. The recovery time is longer than the attack response resulting in a longer sustain and a sonic impression of compression, whereby it is the same way an audio compressor acts. Additionally distortion is likely to appear through voltage drop. This behavior is controlled by resistance and capacitance values in the supply itself [39]. Also aging of components and associated rising of Thevenin resistance, loss of capacity and the number, speed and power of simultaneous keystrokes are a control parameters for this phenomenon colloquially known as "voltage sag". The same is true for either amplification circuits and plate voltage load of capacitive pickups respective microphones. So examinations on electronic components influencing musical parameters are fruitful sources for further studies.

10 Conclusion

In this treatise a fundamental consideration of the tone production mechanisms of the Wurlitzer EP200 series and the Rhodes Mark-II electric pianos was presented. We showed that the characteristic timbre of both instruments is due to the specific setup and geometry of the respective pickup systems. A simplified modeling approach for both instruments was proposed showing good accordance with the measured sounds. Both models are able to run in real-time on a not-so-recent personal computer and can be parametrised for different geometries as well as different pickup designs.

It is hoped-for that this work serves as a starting point for further research regarding the acoustic properties of these or other electro-mechanical instruments. Learning about the fundamental mechanisms of those instruments could help to elucidate the fact why the sound of semi-acoustic instruments are still held in such high regards among listeners and musicians.

Acknowledgments It gives us great pleasure to acknowledge the help of Till Weinreich and Martin Keil who helped performing the acoustic and high-speed camera measurements of the Rhodes E-piano.

[7]The Rhodes sounds on the Billy Cobham's track *Snoopy* is a parade example. The keyboarder, Jan Hammer, uses a ring-modulator to modify the instrument sound.

Appendix I

Constants of the shear beam equation introduced in 13. See: [34]:

k' Shape Factor given as $\frac{6(1+v)}{7+6v}$
v Poisson's Ratio
ρ Dimensionless Density
G Dimensionless Shear Modulus
A Dimensionless Area.

Appendix II

FD approximations can be derived by using the fundamental theorem of calculus, which states that the derivative of a variable function $u(x)$ along dimension x is defined by taking the limit of a *finite* difference Δx of the dependent variable Δu like

$$u_x = \lim_{\Delta x \to 0} \frac{\Delta u}{\Delta x}, \tag{27}$$

with u_x indicating a first derivative by x. For non-zeros but small Δx this expression can be utilized to approximate a differential as a difference

$$u_x \approx \delta_x u \tag{28}$$

with Δ_x indicating a centered first order difference operator by x.

This generalized finite difference operator notation is applied throughout the remainder of this work. It is based on the notation used in works like [30, 40, 41].

A discrete shift operator acting on a 1-dimensional function u at position x is indicated by τ with

$$\begin{aligned} \tau_{x+}(u(t,x)) &= u(t, x+\Delta x), \\ \tau_{x-}(u(t,x)) &= u(t, x-\Delta x). \end{aligned} \tag{29}$$

A difference approximation in the forward (+) and backward (−) direction at position x can be written as

$$\begin{aligned} \delta_{x+}\mathbf{u}|_x &= \frac{1}{\Delta x}(u(x+\Delta x) - u(x)) = \frac{1}{\Delta x}(\tau_{x+} - 1)\mathbf{u}, \\ \delta_{x-}\mathbf{u}|_x &= \frac{1}{\Delta x}(u(x) - u(x-\Delta x)) = \frac{1}{\Delta x}(1 - \tau_{x-})\mathbf{u}. \end{aligned} \tag{30}$$

The same can be done in the temporal domain by defining the forward (+) and backward (−) approximation at time instant t as

$$\begin{aligned}\delta_{t+}\mathbf{u}|_t &= \frac{1}{\Delta t}(u(t+\Delta t)-u(t)) = \frac{1}{\Delta t}(\tau_{t+}-1)\mathbf{u},\\ \delta_{t-}\mathbf{u}|_t &= \frac{1}{\Delta t}(u(t)-u(t-\Delta t)) = \frac{1}{\Delta t}(1-\tau_{t-})\mathbf{u}.\end{aligned} \tag{31}$$

An interesting feature of this operator notation is that higher order approximations can be achieved by a convolution of lower order operators. Using (30), a second order centered finite difference operator can be computed by

$$\begin{aligned}\delta_{xx} &= \delta_{x-} * \delta_{x+}\\ &= [\frac{1}{\Delta x}(1-\tau_{-1})] * [\frac{1}{\Delta x}(\tau_+ - 1)]\\ &= \frac{1}{\Delta x^2}(\tau_+ - 1 - 1 + \tau_-)\\ &= \frac{1}{\Delta x^2}(\tau_- - 2 + \tau_+),\end{aligned} \tag{32}$$

with the equivalence $\tau_+ \cdot \tau_- = 1$.

Higher order operators can be calculated similarly

$$\delta_{4x} = \delta_{xx} * \delta_{xx}. \tag{33}$$

References

1. Shear, G., Wright, M.: The electromagnetically sustained Rhodes piano. In: NIME Proceedings, Oslo (2011)
2. Wendland, Torsten: Klang und Akustik des Fender Rhodes E-Pianos. Technische Universität Berlin, Berlin (2009)
3. Rhodes Keyboard Instruments: Service Manual. CBS musical instruments a division of CBS Inc., Fullerton CA (1979)
4. Smith, R.R.: The History of Rickenbacker Guitars, Anaheim CA (1987)
5. Beauchamp, G.D.: Electrical stringed Musical Instrument, US 2089171 A (1934)
6. Donhauser, Peter: Elektrische Klangmaschinen: die Pionierzeit in Deutschland und Österreich. Böhlau Verlag, Wien (2007)
7. Rhodes, H.B.: Electrical Musical Instrument in the Nature of a Piano, U.S. Patent 2,972,922 (1961)
8. Meyer-Eppler, W.: Elektrische Klangerzeugung: elektronische Musik und synthesische Sprache, Bonn (1949)
9. Rhodes, H.B., Woodyard, S.J.: Tuning Fork Mounting Assembly In Electromechnanical Pianos, U.S. Patent 4,373,418 (1983)
10. Andersen, C.W.: Electronic Piano, US 2974555 (1955)
11. Miessner, B.J.: Method and Apparatus for the Production of Music, US Patent 1,929,027 (1931)

12. Vierling, O.: Das elektroakustische Klavier, Berlin (1936).
13. Vierling, O.: Elektrisches Musikinstrument Electric musical instrument, DE1932V0030615 (1937)
14. Duchossoir, A.R.: Gibson Electrics: The Classic Years: An Illustrated History from the Mid-'30s to the mid-'60s. Hal Leonard Corp., Milwaukee, Wis. (1998)
15. Wheeler, Tom: American Guitars: An Illustrated History. Harper, New York (1992)
16. Palkovic, M.: Wurlitzer of Cincinnati: The Name That Means Music To Millions, Charleston (2015)
17. Espinola, S.: Wurlitzer Electric Piano models: a list.|Paleophone. Blog entry Available at: http://paleophone.net/?page_id=923. Accessed 23rd May 2016
18. Rhodes, H.B.: Piano Action, U.S. Patent 4,338,848 (1982)
19. Wurlitzer Company: The Electric Pianos Series 200 and 200A Service Manual. Available at: http://manuals.fdiskc.com/flat/Wurlitzer%20Series%20200%20Service%20Manual.pdf. Accessed 17 Feb 2016
20. Furukawa, T., Tanaka, H., Itoh, H., Fukumoto, H., Ohchi, N.: Dynamic electromagnetic analysis of guitar pickup aided by COMSOL multiphysics. In: Proceedings of the COMSOL Conference Tokyo 2012, Tokyo. COMSOL (2012)
21. Jian-Ming, J.: The Finite Element Method in Electromagnetics, 3rd edn. Wiley, Hoboken. New Jersey (2014)
22. Martin, M.V.: Finite element modelling of the magnetic field of guitar pickups with ANSYS. Master. University of Applied Sciences, Regensburg (2003)
23. Jin, J.: The Finite Element Method in Electromagnetics, 2nd edn. Wiley-IEEE Press, New York (2002)
24. Jackson, J.D.: Classical Electrodynamics, 3rd edn. Wiley, New York (1998)
25. Furukawa, T., Tanaka, H.: Dynamic electromagnetic analysis of guitar pickup aided by COMSOL Multiphysics, Tokyo (2012)
26. Ippolito, A.C.: Electronic piano feedback reduction, US 3435122 A (1965)
27. Avanzini, F., Rocchesso, D.: Physical modeling of impacts: theory and experiments on contact time and spectral centroid. In: Proceedings of the Conference on Sound and Music Computing, pp 287–293 (2004)
28. Stulov, Anatoli: Hysteretic model of the grand piano hammer felt. J. Acoust. Soc. Am. **97**(4), 2577–2585 (1995)
29. Hunt, K.H., Crossley, F.R.E.: Coefficient of restitution interpreted as damping in vibroimpact. J. Appl. Mech. **42**(2), 440–445 (1975)
30. Bilbao, Stefan D.: Numerical Sound Synthesis: Finite Difference Schemes and Simulation in Musical Acoustics. Wiley, Chichester (2009)
31. Pfeifle, F.: Multisymplectic pseudo-spectral finite difference methods for physical models of musical instruments. In: Bader, R. (ed.) Sound—Perception—Performance, volume 1 of Current Research in Systematic Musicology, pp. 351–365. Springer International Publishing (2013)
32. Pfeifle, Florian, Bader, Rolf: Real-time finite-difference method physical modeling of musical instruments using field-programmable gate array hardware. J. Audio Eng. Soc **63**(12), 1001–1016 (2016)
33. Traill-Nash, R.W., Collar, A.R.: the effects of shear flexibility and rotatory inertia on the bending vibrations of beams. Q. J. Mech. Appl. Math. **6**(2), 186–222 (1953)
34. Han, Seon M., Benaroya, Haym, Wei, Timothy: dynamics of transversely vibrating beams using four engineering theories. J. Sound Vib. **225**(5), 935–988 (1999)
35. Horton, Nicholas G., Moore, Thomas R.: Modeling the magnetic pickup of an electric guitar. Am. J. Phys. **77**(2), 144 (2009)
36. Muenster, M., Pfeifle, F.: Non-linear behaviour in sound production of the Rhodes piano. In: Proceedings of the International Symposium of Musical Acoustics (ISMA) 2014, pp. 247–252, Le Mans, France (2014)

37. Muenster, Malte, Pfeifle, Florian, Weinrich, Till, Keil, Martin: Nonlinearities and self-organization in the sound production of the Rhodes piano. J. Acoust. Soc. Am. **136**(4), 2164–2164 (2014)
38. Blencowe, M.: Designing power supplies for tube amplifiers (2010)
39. O'Connor, K.: The Ultimate Tone, vol. 4, Thunderbay (2006). (Chaps. 4–5)
40. Jordan, Charles: Calculus of finite differences, 1st edn. Chelsea Publishing Co, New York (1950)
41. Strikwerda, J.C.: Finite Difference Schemes and Partial Differential Equations, 2nd edn. Society for Industrial and Applied Mathematics, Philadelphia, USA (2004)

Author Biographies

Florian Pfeifle received his M.A. in systematic musicology and electrical engineering in 2010 and his Ph.D. in 2014. His current research is concerned with real-time grand piano physical modeling on FPGAs. Additionally, he is interested in instrument acoustics, scientific computation, and jazz music research. Until recently, he worked also as a musician/producer being awarded gold records for two of his works.

Malte Münster studied musicology, psychology, and phonetics. In addition, he studied piano, guitar and composition. He holds a Master's degree in Systematic Musicology and currently completes a Ph.D.-thesis in the field of musical acoustics with a focus on bronze metallophones. He did ethnomusicological fieldwork in Brazil and in areas of the Balkan Peninsula where he recorded a wide range of rural styles. Malte Münster is also active as an engineer and producer recording various artists in his studio.

Feedback of Different Room Geometries on the Sound Generation and Sound Radiation of an Organ Pipe

Jost Leonhardt Fischer

The trick is the idealizations.
Richard Feynman, Lectures on Physics, Volume I, Chap. 12.1

Abstract Feedback effects of different room geometries on the sound generation and sound radiation of an organ pipe is discussed. Motivation of the present work is that in real organs many of the organ pipes cannot radiate sound without disturbance. Most of the organ pipes are mounted, concealed from the audience's view, behind the organ's prospect. Organ pipes of the same stop and with nearly identical timbre are arranged closely on the wind-chest. With several pipe ranks, sound radiated is reflected from pipes as well as from structural elements such as beams or mounting brackets, which hinder free sound radiation. The investigations were carried out by numerical simulations of an organ pipe and its aeroacoustical interaction with several principally different wall geometries as well as swell chambers. The investigation focuses on the effects of reflected sound waves upon the sound generator region of the organ pipe. Several general set-ups were implemented in a numerical space, a pseudo-3*D* computational grid. The numerical simulations were calculated by solving the fully compressible Navier-Stokes equations with suitable initial and boundary conditions for different geometric constraints using a proper LES-model.

1 Introduction

In this chapter, the feedback effect of different room geometries on the sound generation and sound radiation of an organ pipe is discussed. The motivation for the present work is the fact that in real organs most of the organ pipes are mounted, concealed from the audience's view, behind the organ's prospect [1]. Pipes of the same organ stop are arranged closely on the wind-chest [1] so that the sound from each pipe cannot radiate without disturbance. Reflections on beams, mounting

J.L. Fischer (✉)
University of Hamburg, Hamburg, Germany
e-mail: jost.leonhardt.fischer@uni-hamburg.de

A. Schneider (ed.), *Studies in Musical Acoustics and Psychoacoustics*,
Current Research in Systematic Musicology 4,
DOI 10.1007/978-3-319-47292-8_4

brackets and other structural elements hinder the free radiation of sound and may influence the sound generation and radiation significantly.

The investigations were carried out by numerical simulations of an organ pipe and its aeroacoustical interaction with several principally different wall geometries. The investigations are focused on the effects of reflected sound waves upon the sound generator region of the organ pipe. The original organ pipe used as template for the numerical model was made by the German organ builder Alexander Schuke Orgelbau Potsdam GmbH [2], cf. Fig. 1a, b.

Several general set-ups were implemented in a numerical space, a pseudo-3*D* computational grid. The numerical simulations were calculated by solving the fully compressible Navier-Stokes equations with suitable initial and boundary conditions for different geometric constraints using a proper LES-model.

The numerical simulations were realized by using parts of the C++ toolbox OpenFoam-2.1. [3]. These libraries include customized numerical solvers as well as pre- and post-processing utilities for the solution of continuum mechanics problems, including computational fluid dynamics (CFD) and computational aeroacoustics (CAA). The code is released as free and open source software under the GNU General Public License. OpenFOAM stands for Open source Field Operation And Manipulation. For details regarding implementation, run and post-processing techniques the reader is referred to the relevant OpenFOAM User Guide and the OpenFOAM Programmer Guide.

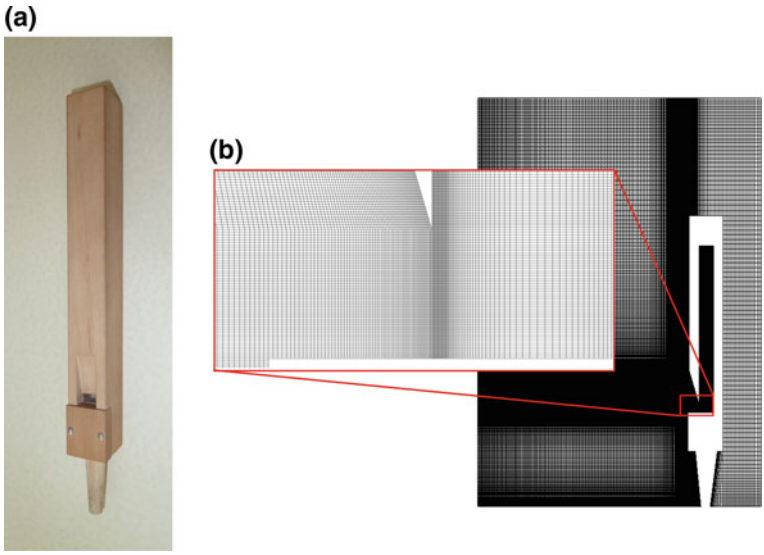


Fig. 1 **a** Wooden organ pipe with quadratic cross-section, built and provided for measurement use by organ builders Schuke Orgelbau Potsdam GmbH. **b** Implementation of the organ pipe and the surrounding space into a pseudo-3D computational grid. The detail gives an impression of the mesh size in the cut-up region

The first section gives some notes about the general procedure of successful implementation and run of such advanced numerical simulations.

In the second section, the impact of various room geometries on the sound generation of the organ pipe is analyzed at the basis of several particular wall geometries, planar and non-planar geometries. The analyzed wall geometries are representative of conditions that frequently occur in real pipe organ buildings.

The third section addresses the influence of a planar, nearly-closed box geometry, emulating the acoustical conditions of an organ pipe mounted within an open-backed swell chamber. Swell chambers are an integral component of many pipe organs. The organ pipes are mounted within a wooden box equipped with movable shutters towards one, or several, sides which may be mechanically regulated (opened or closed). The swell chamber allows for adding increased dynamic expressiveness to the stop's otherwise rigid behavior, once the instrument maker has finished the intonation process. Examined herein is the influence of the swell chamber geometry on both timbre and frequency spectrum of the organ pipe.

2 General Notes on Numerical Implementation and Numerical Simulation

The sound generation in organ pipes, the interaction between flow field (wind field) and acoustical field as well as the sound propagation are described by the compressible Navier-Stokes equations [4, 5]. Hence the compressible Navier-Stokes equations have to be solved with given initial and boundary conditions on a corresponding computational grid, the numerical space, called mesh.

The numerical treatment of compressible problems is an advanced task. In general, a successful procedure of realization can be divided into four sections. (O) Physical previews (A) Pre-processing, (B) Processing and (C) Post-processing. The sections include the following sub-tasks and questions being answered:

O 1. What set of equations describes the problem?
O 2. What characteristic fluid dynamical numbers one has to take into account?
O 3. What are the scales of the problem?
O 4. Hardware-decision
O 5. Software-decision

A 1. How to program a proper mesh?
A 2. Which relevant thermo-physical properties have to be configured?
A 3. Implementation of suitable initial and boundary conditions for physical quantities, e.g. pressure p, velocity vector $\vec{U}$, temperature T, density ρ, turbulent kinetic Energy k, etc.
A 4. Appropriate discretization schemes for the operators (Del-operator, Laplacian, time derivative, etc.), inclusive proper correctors

A 5. Identification of an appropriate turbulence model to model the energy transport into the sub-grid scales
A 6. Adequate solver for compressible fluid dynamical problems, determination of numerical schemes and their tolerances
A 7. Adequate matrix solvers
A 8. Configuration of numerical parameter, e.g. numerical time step size, simulation time, write precision etc.
A 9. Definition of suitable probe points and (or) suitable sample sets in the mesh for analysis
B 1. Parallelization of the simulation
B 2. Numerical stability parameter, e.g. Courant number
B 3. Control during simulation run time
B 4. Calculation of additional physical quantities from the data

C 1. Visualization
C 2. Analysis.

For more information, the reader is referred to the author's Ph.D. thesis [6]. General aspects about implementation, pre-processing, run and post-processing numerical simulations can be found in the OpenFOAM User Guide as well as the OpenFOAM Programmer Guide [3].

3 The Effect of Complex Geometries

To study the organ sound affected by different spatial geometries, the following scenarios are considered:

scenario: *wall*	planar wall at distance 140 mm,
scenario: *wall_lambda*	planar wall at distance $\frac{\lambda}{4} = 125$ mm,
scenario: *convex*	convex wall,
scenario: *concave*	concave wall,
scenario: *diffuse*	ridged wall,
scenario: *free*	no walls.

The scenarios were transferred into respective pseudo-3D computational grids. Pseudo-3D means that the mesh besides the x- and y-dimension has also a z-dimension of 10 mm but with only one face in depth. The meshes of the specified scenarios are shown in Fig. 2. The walls considered acoustically inert are marked red. The scenario *free*, which was used as reference scenario has no walls at the boundaries of the numerical set-up. This means that the radiated sound can propagate through the boundaries without any restrictions. The key data of the generated meshes are listed in Table 1. The technique how to write a proper mesh file and how to generate a mesh is documented in the OpenFOAM User Guide [3].

Other configurations, e.g. the initial conditions for the physical quantities velocity $\vec{U}$, pressure p, temperature T, density ρ, turbulent, kinetic energy k, etc., thermo-physical

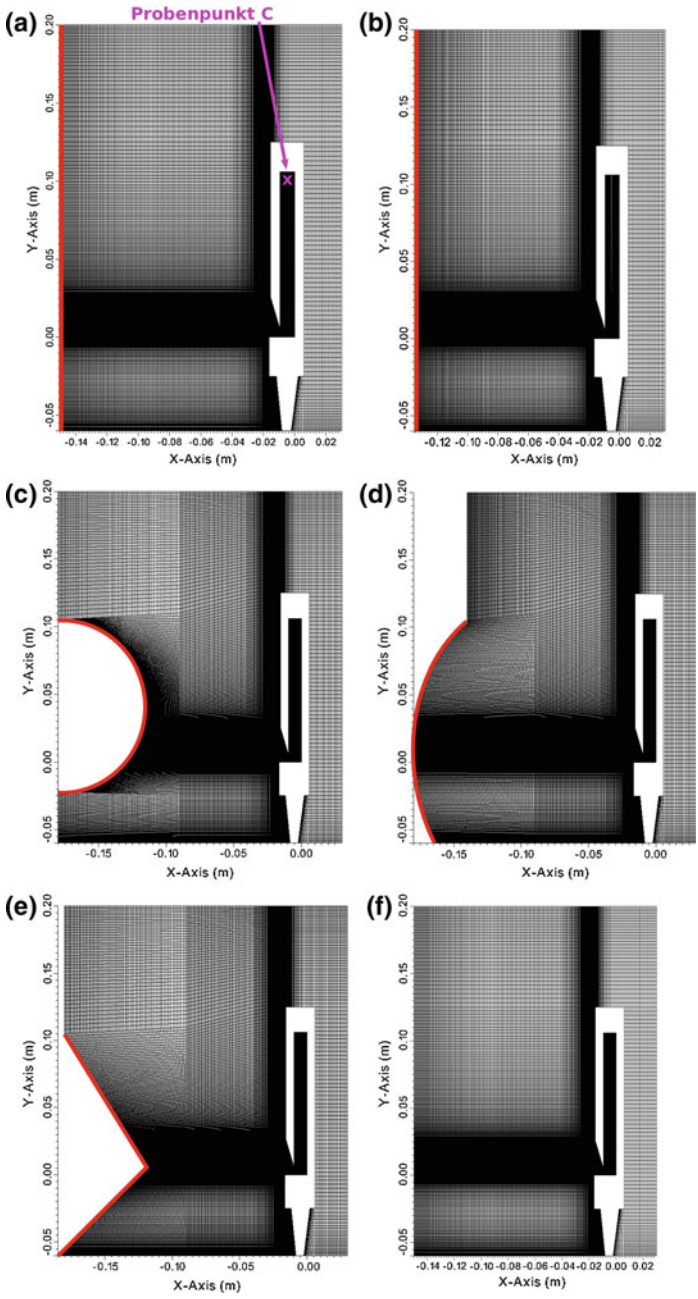


Fig. 2 The meshes of the scenarios which were investigated. The acoustically inert walls are marked *red*: **a** plane wall (*wall*) at distance 140 mm, **b** plane wall at distance $\lambda/4 = 125$ mm relative to the mouth (*wall_lambda*), **c** convex wall (*convex*), **d** concave wall (*concave*), **e** ridged wall (*diffuse*), **f** free space, without walls (*free*) utilized as reference scenario. The analysis of the scenarios refer to the probe point *C*, which is marked in the mesh of scenario *wall*

Table 1 Key data of the meshes of the different scenarios

Scenario	Mesh points	Faces	Hexaeders
free	254342	505170	126000
wall	254342	505170	126000
wall_lambda	260362	517180	129000
convex	260362	517180	129000
concave	260362	517180	129000
diffuse	260362	517180	129000

Table 2 Thermo-physical properties of air at temperature $T = 20$ °C and normal air pressure

Property	Value	Unit
Molecules	1	
Molar mass M	28.9	kg/kmol
Thermal capacity ($p = const$) C_p	1007	J/kg/K
Latent heat h	0	(off)
Dynamic viscosity μ	1.8×10^{-5}	Ns/m^2
Prandtl number Pr	0.7	–

properties, like molar mass M for the medium air, which is assumed as a perfect gas, heat capacity at constant pressure C_p, latent heat h, dynamic viscosity μ, Prandtl number Pr as well as turbulence properties were implemented. A suitable LES-Model, namely a dynamic SGS-k-model, was choosen. The LES-model includes a transport equation for the turbulent kinetic energy k to calculate the energy transfer into the sub-grid scales (SGS). The thermo-physical properties are summarized in Table 2.

More information about how to implement the mentioned properties can be found in the OpenFOAM User Guide [3]. Visualizations of the numerical simulations are done by the open-source, multi-platform data analysis and visualization application ParaView. Exemplary shown are sequences of the scenario *free* for the quantities pressure p (Fig. 3) and turbulent kinetic energy k, (Fig. 4) as well as a sequence of scenario *wall* of the quantity velocity magnitude $|U|$, (Fig. 5).

The analysis mainly refers to the investigation of the pressure at the sample point C. The probe point C in all scenarios is 10 mm, centrally located beneath the closed resonator end (cf. Fig. 2). At first, the initial excitation process, called the initial excitation process, is investigated. Then the sound pressure signals of the various scenarios are compared and the sound pressure level spectra (SPL-spectra) are investigated and compared with the corresponding signal of the reference scenario *free*.

3.1 The Initial Excitation Process

During the initial excitation process, the organ pipe ‘finds’ its sound. From a fluid mechanical perspective, in this timespan the initial coupling of the wind field of the inflow and the acoustic field occurs, with the result that in the optimal case a periodic oscillating air sheet is formed, called the jet. The periodical oscillations are

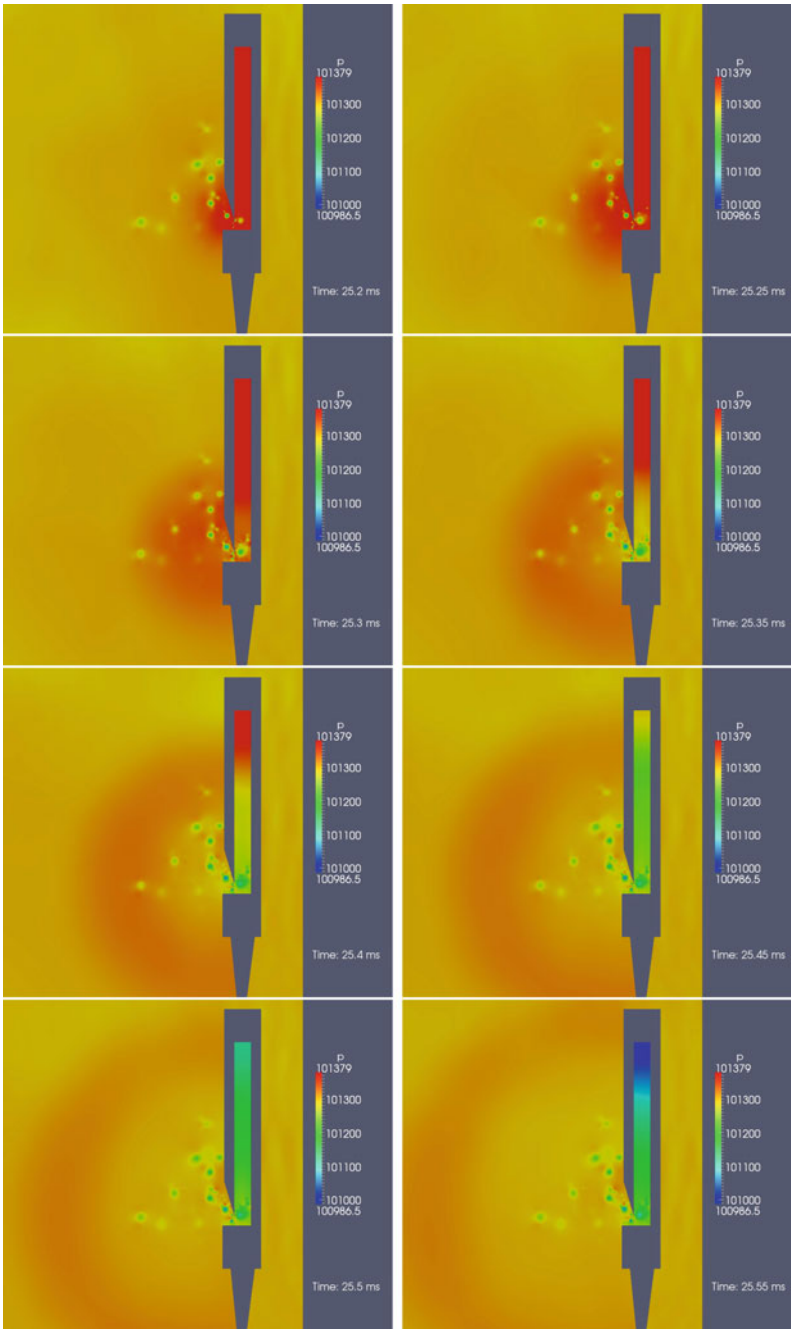


Fig. 3 Sequence $t = 25.2 - 25.55$ ms of the numerical simulation of scenario *free*. Shown is the pressure p. Depicted is the radiation of a sound wave into free space

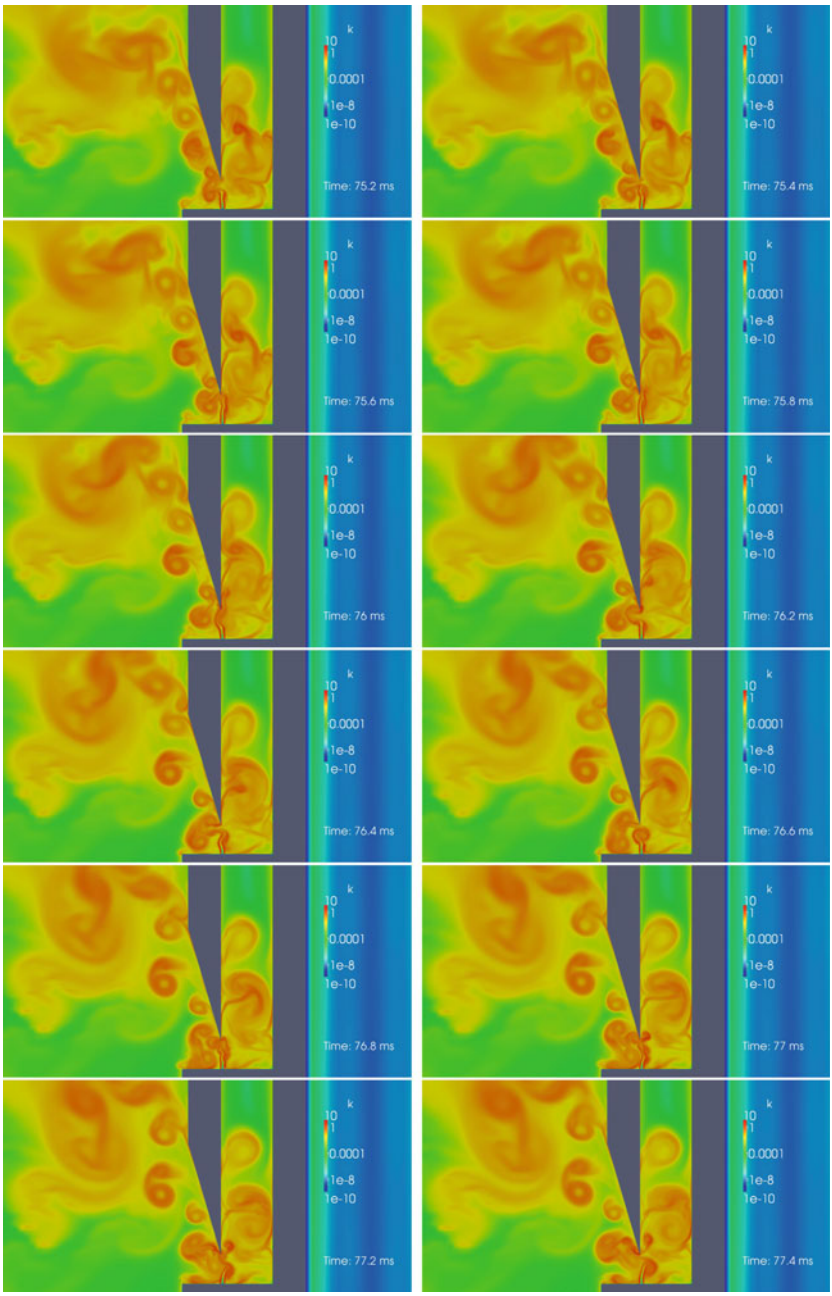


Fig. 4 Sequence $t = 75.2 - 77.4$ ms of the numerical simulation of scenario *free*. Shown is the log-scaled turbulent kinetic energy k. Depicted is the oscillating jet with its shear layers, the formation of vortices inside the resonator as well as a vortex street, escaping the organ pipe along the upper labium

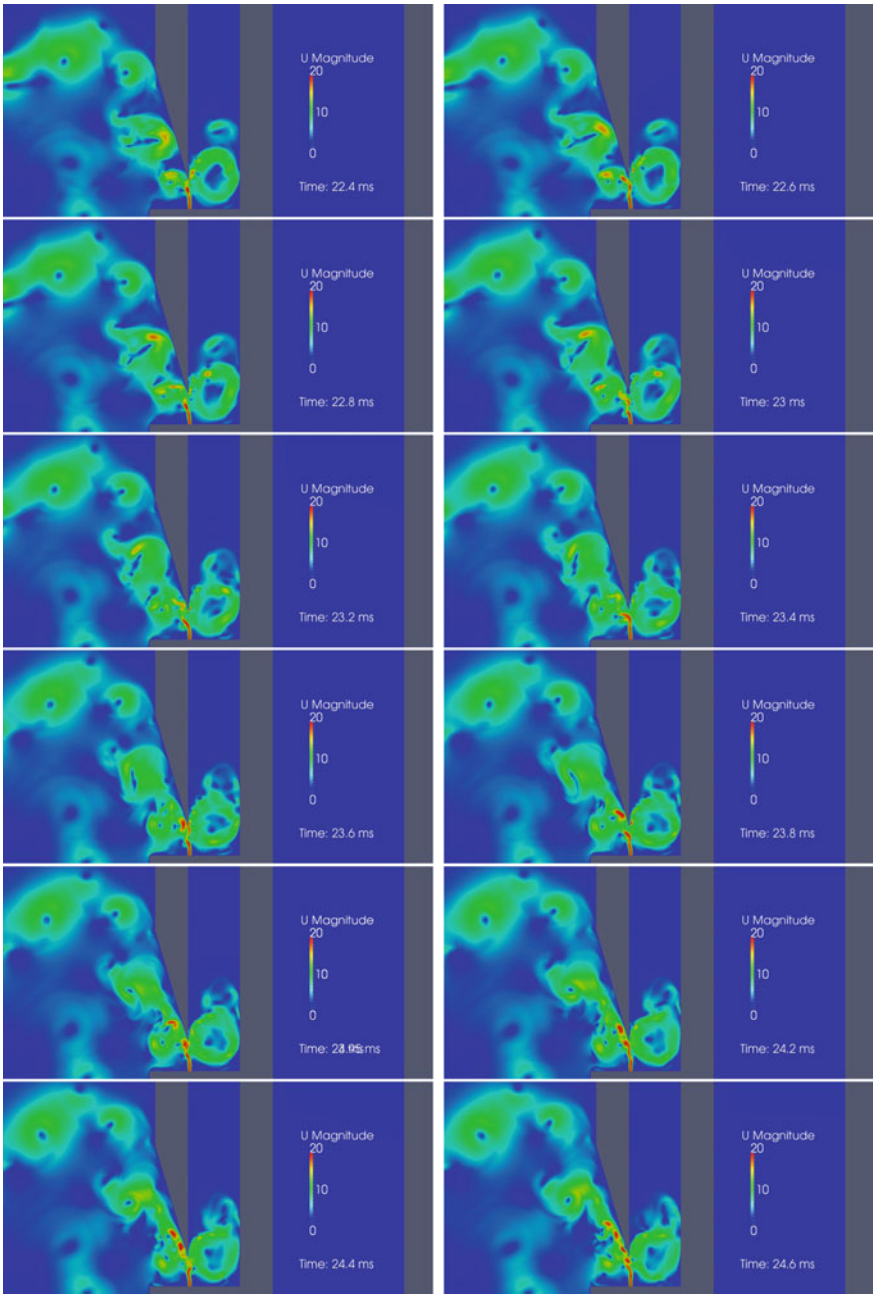


Fig. 5 Sequence $t = 22.4 - 24.6$ ms of the numerical simulation of scenario *free*. Shown is the velocity magnitude $|U|$. Depicted is the oscillating jet with its shear layers, the formation of vortices inside the resonator as well as a vortex street, escaping the organ pipe along the upper labium

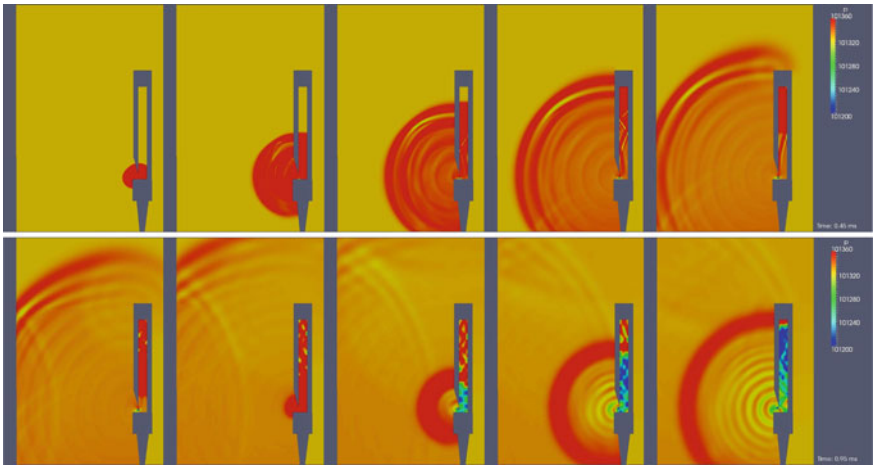


Fig. 6 Visualization of the radiation of the initial sound waves of a stopped organ pipe in the reference scenario *free*

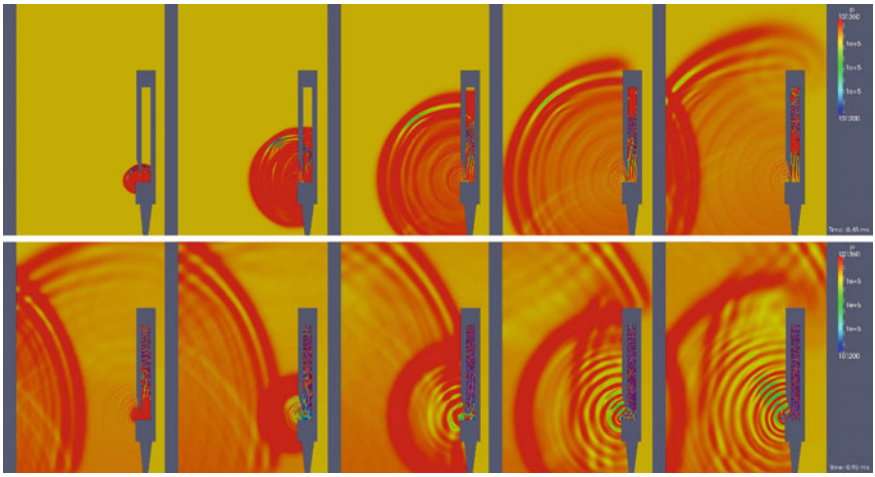


Fig. 7 Visualization of the radiation of the initial sound waves of a stopped organ pipe in the reference scenario *wall*

transverse to the main direction of the flow. The transient process of initial excitation is dependent on a large number of physical and geometrical parameters such as the wind speed, the size and position of the orifice and the geometry of the mouth. The geometry of the resonator, as well as the geometry of the surrounding room, is this study's object of investigation. The duration of the initial process of excitation lay, in case of the organ pipe examined, in the range of the first 10 ms. Figures 6, 7, 8, 9, 10 and 11 are examples of sequences of the initial excitation for

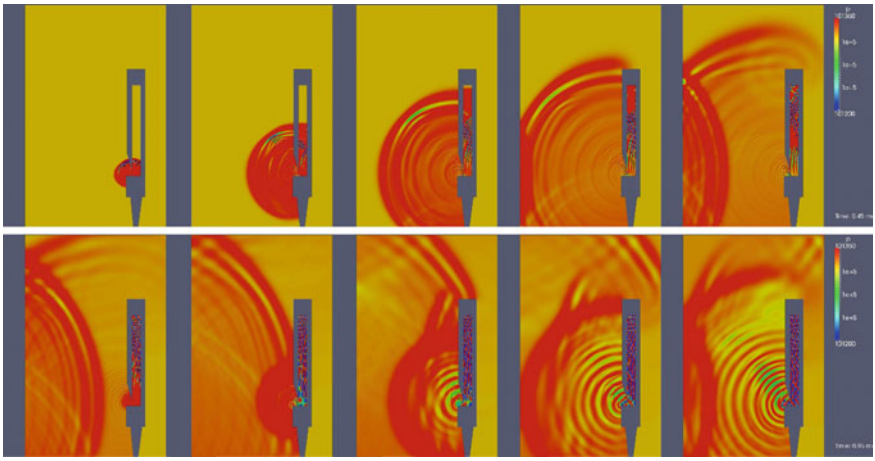


Fig. 8 Visualization of the radiation of the initial sound waves of a stopped organ pipe in the reference scenario *wall_lambda*

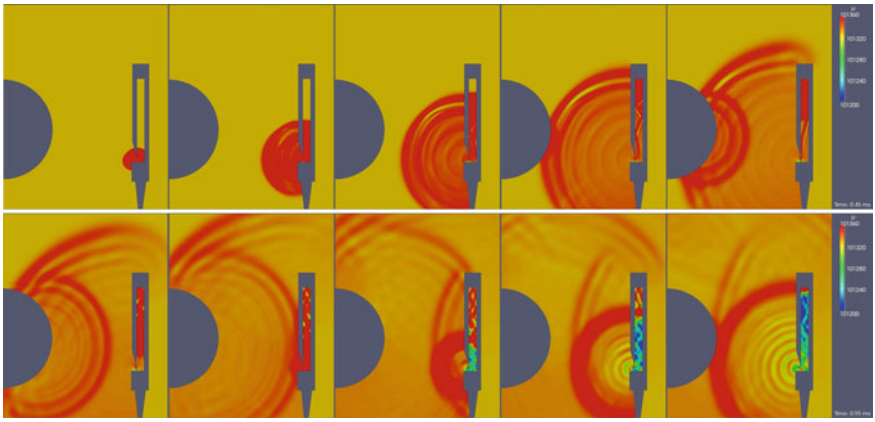


Fig. 9 Visualization of the radiation of the initial sound waves of a stopped organ pipe in the reference scenario *convex*

scenarios *free*, *wall*, *wall_lambda*, *convex*, *concave* and *diffuse*. It can be seen how the initial sound pressure wave is generated and subsequently radiated into the surrounding room.

The reference scenario *free* shows that sound generation and sound radiation are well represented by the numerical simulation. The propagation of the initial sound pressure wave inside the resonator is clearly shown. Also the transverse modes typical for the initial excitation process are discernable. Because of the pseudo-3D computational grid, the sound pressure wave radiating into the surrounding room appears as a circular shape. For the three-dimensional case, spherical waves are

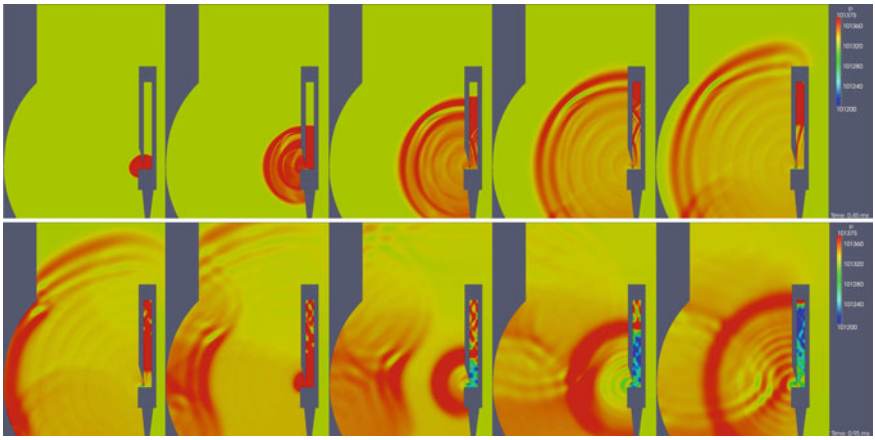


Fig. 10 Visualization of the radiation of the initial sound waves of a stopped organ pipe in the reference scenario *concave*

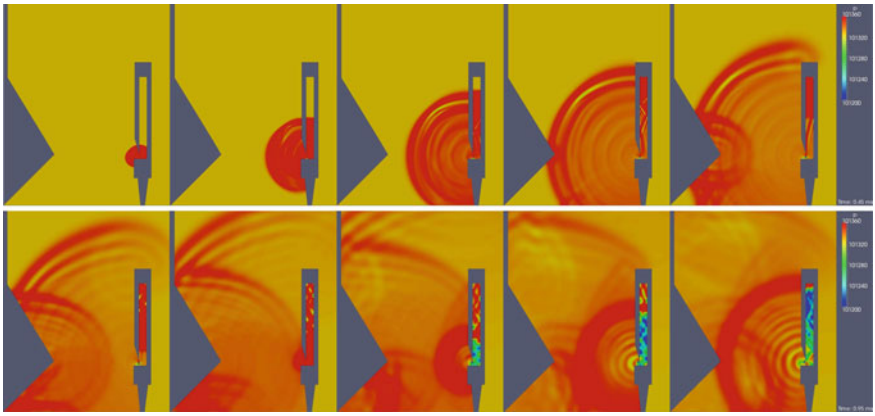


Fig. 11 Visualization of the radiation of the initial sound waves of a stopped organ pipe in the reference scenario *diffuse*

assumed. Also visible is the radiation of higher harmonics. The scenario's outer margins present a particular challenge, as the boundary conditions have to be defined such that an unimpeded transmission of the propagating physical values is ensured. This means that, within one mesh width, either complete absorption has to be assumed as the boundary condition, this often leading to strong numerical reflections, or a more realistic boundary condition, allowing for the propagation of waves beyond the limits of the computational grid, has to be chosen. Here, the latter variant has been implemented. In the initial conditions for the pressure, respective far field conditions have been specified. At the computational grid's margins, therefore, only comparatively minor numerical reflections occur. They appear at the

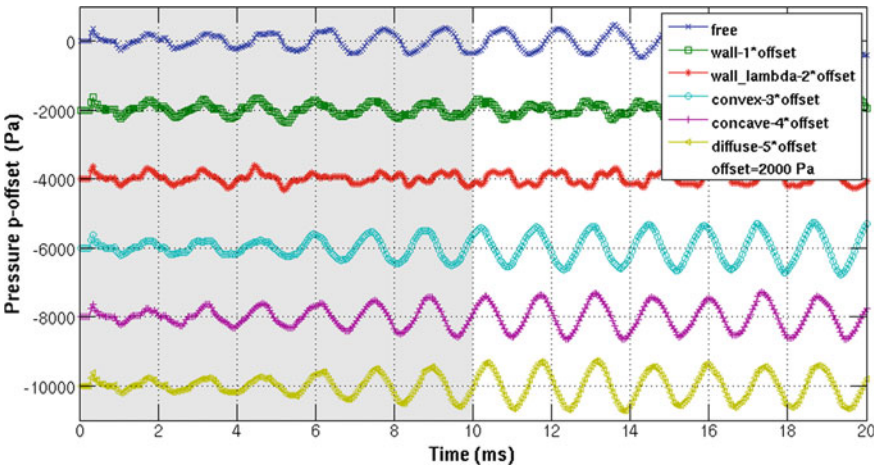


Fig. 12 The initial excitation process. Shown are the pressure signals of the scenarios sampled at probe point *C*

order of magnitude of the acoustical diffraction phenomena at the upper end of the organ pipe, observable as a bright circle in the first illustration of the second sequence row ($t = 0.55$ ms). Reducing these numerical reflections further by fine tuning the transmission boundary conditions is the aim of current and future research.

Figures 7 and 8 depict sequences of the initial excitation processes of scenarios *wall* and *wall_lambda*. In comparison to the reference scenario *free*, these scenarios show the reflection of the sound wave at the acoustically inert walls. The reflected sound wave propagates back towards the cut-up region of the organ pipe. Observing the radiated higher harmonics in the wake of the initial sound wave, superpositions are clearly visible. Also, the pronounced transverse modes inside the resonator indicate that the reflected sound wave seriously interferes with the periodical movements of the jet. The sound wave is reflected once more at the outer surface of the organ pipe. The space between organ pipe and wall thus becomes a kind of "outer resonator", with the length of the resonator being the distance between wall and cut-up.

The sequence of the initial excitation process of scenario *convex* is shown in Fig. 9. This sequence illustrates how the radiated initial sound wave is reflected at the boundary's convex geometry, propagating back towards the cut-up region as a spherical wave. Note that the radius of the reflected sound wave is smaller than the radii in the scenarios with planar walls. In the sequence of scenario *concave* it can be seen how the initial sound pressure wave is reflected at the boundary's concave geometry, being subsequently focused. In the sequence of scenario *diffuse*, the initial sound pressure wave is reflected at the boundary's ridged geometry, so that the reflected wave is separated.

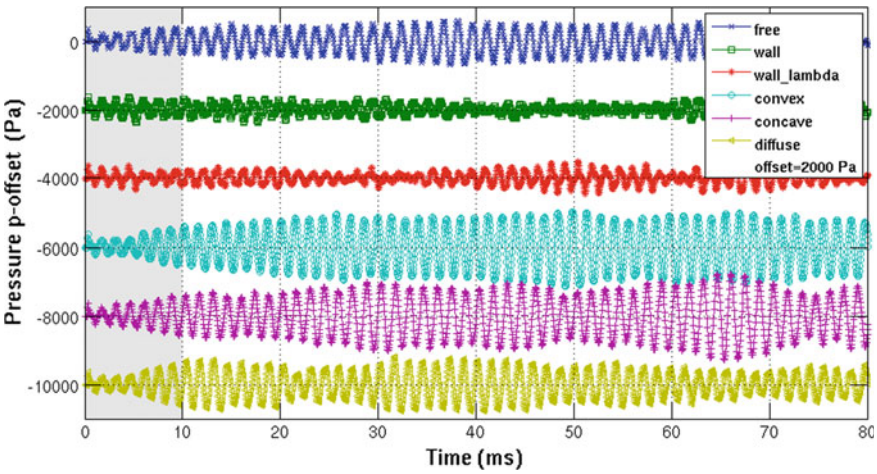


Fig. 13 The pressure signals sampled at probe point *C* for the different scenarios

In Fig. 12, the sound pressure signals during the initial excitation process are plotted at probe point *C* for all computed scenarios. The initial excitation process is shaded in gray. For reasons of clarity, the signals have been separated by an offset of -2 kPa. In the reference scenario *free*, regular, periodic oscillations develop after approximately 10 ms.

In the scenarios *wall* and *wall_lambda* the formation of periodic oscillations is seriously impaired. The disturbances cause a decrease in amplitude as well as a doubling of period. The signals of scenarios *convex*, *concave*, and *diffuse*, on the other hand, exhibit no discernable impairment of the initial excitation process. To the contrary, these scenarios are characterized by a smooth and consistent transient behavior. In comparison to the reference scenario, the developing amplitudes are larger. The oscillations show a triangular waveform, indicating odd-numbered frequencies. Relative to the reference, scenario *diffuse* exhibits a phase shift of $\pi/2$ at $t = 10$ mm. Consequently, the fundamental frequency of the system is higher than in the reference scenario.

The signals resulting from the further course of simulations are shown in Fig. 13. The excitation process again is shaded in gray. The disturbance of the jet in scenarios *wall* and *wall_lambda* leads to a decrease in amplitude as well as to beating. In scenario *wall* the beating exhibits a period of approx. $T_S = 100$ ms, corresponding to beat frequencies of approx. $f_S = 10.0$ Hz and $f_S = 16.5$ Hz, respectively. Beating can only be induced by the superposition of several different frequencies. This implies that a plane wall acts as a significant obstacle to sound radiation, causing attenuation in the amplitude domain as well as amplification in the frequency domain.

The pressure signal in scenario *convex*, at sample point *C*, exhibits the most regular waveform of all scenarios, including the reference scenario *free*. Amplitudes are consistently larger compared to the reference case. Furthermore, a slight

increase in frequency is observed. The results show that this geometry leads to an amplification in both the amplitude and frequency domains.

In scenario *concave*, the largest amplitudes are observed. This is caused by the wall's concave geometry focussing the reflected sound wave. The amplitudes of the signal are, however, less evenly distributed when compared with *convex*, perhaps also due to beating.

The waveform of scenario *diffuse* is also uneven and exhibits beating. Also, the separation and diversion of the reflected sound wave causes lower amplitudes.

3.2 Sound Pressure Level Spectra

Based on the pressure signals at probe point C, as obtained by numerical simulation, sound pressure level spectra are generated. Here the signals obtained over the simulation's entire duration of $t_{sim} = 100\,\text{ms}$ are utilized. The sampling frequency is the inverse of the sampling interval. In the numerical simulations, this is the temporal increment displayed as $\Delta t_s = 5 \times 10^{-6}\,\text{s}$. This results in a sampling frequency of $f_s = 20000\,\text{Hz}$.

First, the extracted signals are transformed into amplitude spectra by means of Fourier transformation. The concept of Fourier transformation constitutes the notion that any periodic function with a period T can be expressed as the sum of a, generally infinite, number of harmonic oscillations with their respective specific frequencies being integral multiples of the fundamental frequency f_0. In the case of the function being a discrete signal, the decomposition by Fourier is, assuming $\omega_0 = 2\pi/T = 2\pi f_0$, expressed as

$$f(t) = c_0 + \left[\sum_{n=1}^{\infty} a_n \cdot \cos(n\omega_o t) + b_n \cdot \sin(n\omega_0 t)\right] \tag{1}$$

The coefficients $a_n(n\omega_0)$ and $b_n(n\omega_0)$ are called the Fourier coefficients. To determine the Fourier coefficients is the subject matter of Fourier analysis. One finds

$$c_0 = \frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} f(t)\,dt \tag{2}$$

$$a_n = \frac{2}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} f(t) \cdot \cos(n\omega_0 t)\,dt \quad n = 0, 1, 2, 3, \ldots \tag{3}$$

$$b_n = \frac{2}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} f(t) \cdot \sin(n\omega_0 t)\, dt \quad n = 0, 1, 2, 3, \ldots \tag{4}$$

The Fourier coefficient c_0 is the temporal average of the signal $f(t)$ taken over one period, also referred to as steady component, or as the signal's offset (in electrical engineering called DC-component).

Taking the relations

$$a_n \cos(n\omega_0 t) + b_n \sin(n\omega_0 t) = c_n \sin(n\omega_0 t + \phi_n) \tag{5}$$

and

$$c_n = \sqrt{a_n^2 + b_n^2} \tag{6}$$

$$\phi_n = \arctan\left(\frac{a_n}{b_n}\right) \tag{7}$$

one finally obtains the **spectral representation** of the Fourier series

$$f(\omega) = c_0 + \sum_{n=1}^{\infty} c_n \sin(n\omega_0 t + \phi_n) \tag{8}$$

A periodic signal is thus determined by the values for

c_0	Temporal average over one period of the signal f(t),
$c_n = c_n(n\omega_0)$	Amplitude spectrum,
$\phi_n = \phi_n(n\omega_0)$	Phase spectrum.

Taking the amplitude spectrum, the sound pressure level spectrum (SPL-spectrum) is calculated by

$$\mathrm{SPL} = 20 \cdot \log_{10}\left(\frac{p_{rms}}{p_0}\right) \mathrm{dB} \tag{9}$$

with the pressure's root mean square value $p_{rms} = p/\sqrt{2}$ and the reference pressure being $p_0 = 20\,\mu\mathrm{Pa}$.

Figure 14 shows the sound pressure level spectra obtained. In Tables 3 and 4 the frequencies and amplitudes of the fundamentals as well as of the 2nd, 3rd, and 5th harmonics are summarized. The frequency resolution of the numeric simulation's level spectra result from the inverse of the simulation's duration and amounts to $\Delta f = \pm 5\,\mathrm{Hz}$.

All level spectra exhibit a prominent fundamental oscillation. Compared to the spectrum of reference scenario *free*, the largest increase in frequency is to be found

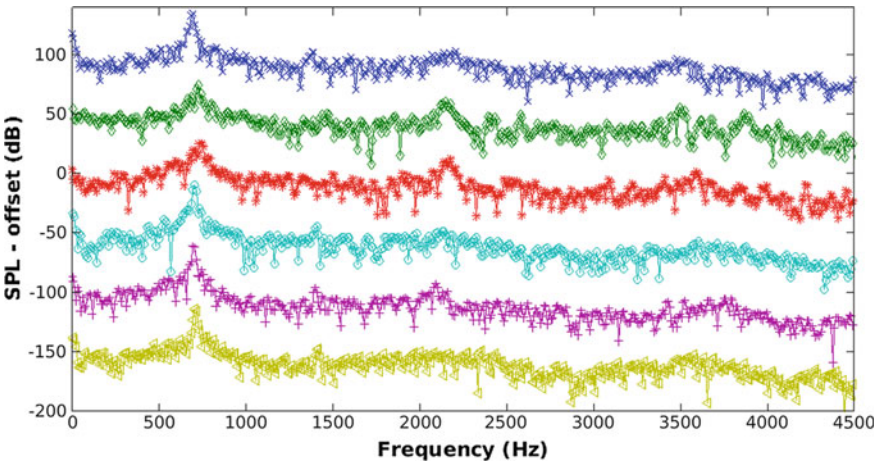


Fig. 14 SPL-spectra at the probe point *C* for the scenarios, top down, *free*, *wall*, *wall_lambda*, *convex*, *concave* and *diffuse*. Shown are the spectra of sound pressure level in the range of frequency of $0 - 4500$ Hz. For reasons of clarity the spectra are separated by an offset of -50 dB. The SPL-scale is related to the reference scenario *free*. One sees prominent fundamentals, 2nd harmonics with low amplitudes, 3rd, and 5th harmonics

Table 3 Frequencies of the fundamentals, the 2nd, 3rd and 5th harmonics sampled at probe point *C*

Frequency/scenario	Fundamental (±5 Hz)	2. Harm. (±5 Hz)	3. Harm. (±5 Hz)	5. Harm. (±5 Hz)
free	693	1377	2207	3477
wall	723	1465	2148	3496
wall_lambda	742	1507	2178	3594
convex	703	1406	2129	3604
concave	693	1396	2090	3750
diffuse	703	1416	2002	3574

Table 4 SPL-spectra of the fundamentals, the 2nd, 3rd, and 5th harmonics sampled at probe point *C*

SPL/scenario	Fundamental (dB)	2. Harm. (dB)	3. Harm. (dB)	5. Harm. (dB)
free	131.5	99.5	100.5	93.5
wall	121.5	97.5	106.5	102.5
wall_lambda	122.5	96.5	109.5	98.5
convex	137.5	101.5	97.5	89.5
concave	136.5	100.5	104.5	92.5
diffuse	132.5	99.5	96.5	94.5

in scenarios *wall* and *wall_lambda*. They amount to $30 \pm 5\,Hz$ and 49 ± 5 Hz, respectively. The level of the 3rd and 5th harmonics, furthermore, are significantly emphasized. One finds increases in level of +6 and +9 dB of the respective frequencies, cf. Table 3.

Scenario *convex* is distinguished by the accentuation of the fundamental oscillation when compared to the reference scenario. The level is raised by $+6$ dB as against scenario *free*, meaning a doubling of the fundamental oscillation's loudness. All other harmonics, on the other hand, are slightly attenuated.

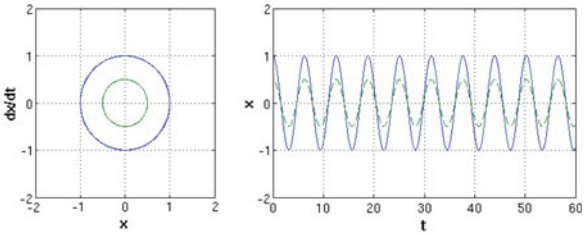
In scenario *concave*, there is an increase in level of the fundamental as well as the 3rd harmonic, caused by the concave geometry. The ridged geometry in scenario *diffuse* accentuates the fundamental and attenuates the 3rd harmonic.

In all scenarios, the 2nd harmonic can be observed. The occurrence of the second harmonic in a stopped organ pipe is, admittedly, mathematically impossible according to linear wave theory. Organ builders, however, are familiar with this effect in their practical work. The cause for the occurrence of the 2nd harmonic is the formation of an oscillating acoustic dipole at the tip of the labium [7]. The dipole changes its configuration twice within each of the jet's periods, i.e., it oscillates with twice the jet's frequency. The level of the 2nd harmonics observed in the numerical simulation are usually smaller than the levels of the fundamentals and the 3rd harmonics; this is in agreement with practical experience. The attenuation of higher harmonics as observed in scenarios *convex* and *diffuse* leads, however, to the 2nd harmonic obtaining a certain sonic significance. The occurrence of 2nd harmonics in the numerical simulation attests for its high accuracy and realism.

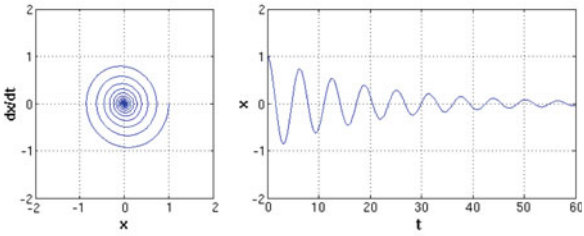
3.3 Phase Portraits

Taking the time derivative of the scenarios' pressure signals, and plotting dp/dt against p, the results are the phase portraits of the oscillating pressure at probe point C (cf. [7]). The phase portraits provide information about the similarity of the observed oscillations to known types of oscillators. An harmonic oscillator, for example, displays a closed circular trajectory; a damped harmonic oscillator with positive damping displays a spiral shape wound around a stable central fixpoint; an excited harmonic oscillator, having negative damping, displays a spiral shape winding itself away from an instable central fixpoint. A self-excited oscillator, a resonator with nonlinear damping, displays a limit cycle. Depending on the degree of nonlinearity, its phase portrait exhibits a more or less deformed circle. Figure 15a–d as exemplary models show the phase portraits of different oscillator types.

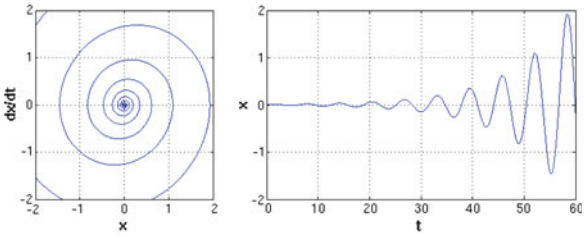
In Fig. 16a–f, the smoothed phase portraits of the different scenarios within the range t = 10–80 ms of the numerical simulations are shown. Figure 17a–f show the normalized phase portraits. The transient excitation process is truncated. Highly dissimilar trajectories can be seen. With the signals being constantly subjected to external interferences, no cohesive trajectories, as described in theory, are observed. The scenarios indeed represent extensive aeroacoustic, and in a sense 'real',



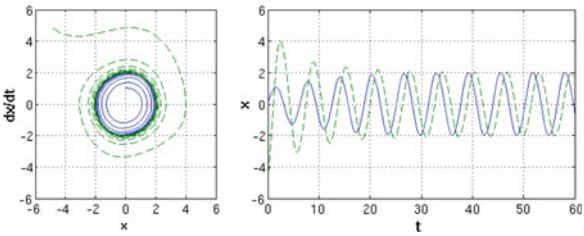
(a) The harmonic oscillator: $\ddot{x} + \omega_0^2 x = 0, \omega_0 = 1.0$. Shown are two solutions for the initial conditions: $x_1(0) = 1.0, \dot{x}_1 = 0.0$, and $x_2(0) = 0.5, \dot{x}_2 = 0.0$.



(b) A damped harmonic oscillator:
$\ddot{x} + \gamma\dot{x} + \omega_0^2 x = 0, \omega_0 = 1.0, \gamma = 0.1$. The initial conditions are:
$x(0) = 1.0, \dot{x} = 0.0$



(c) A negatively damped harmonic oscillator:
$\ddot{x} + \gamma\dot{x} + \omega_0^2 x = 0$, $\omega_0 = 1.0, \gamma = -0.18$. The initial conditions are:
$x(0) = 0.0, \dot{x} = 0.01$



(d) A self-sustained oscillator type, the Van der Pol oscillator:
$\ddot{x} - \mu\left(1 - x^2\right)\dot{x} + \omega_0^2 = 0$, $\omega_0 = 1.0, \mu = 0.08$. The initial conditions of the two solutions are:
$x_1(0) = 0.0, \dot{x}_1 = 1.0, x_2(0) = -5.0, \dot{x}_2 = 5.0$

Fig. 15 **a** The harmonic oscillator as an example of a conservative system. The total energy of the system is constant. **b** The damped harmonic oscillator, a dissipative system. Energy gets lost. **c** The harmonic oscillator with a negative damping term, an (internal) excitative system. Energy is supplied. **d** The Van der Pol oscillator as an example of a self-sustained oscillator. The system balances dissipation and internal excitation. After a transient timespan the system oscillates at the limit cycle

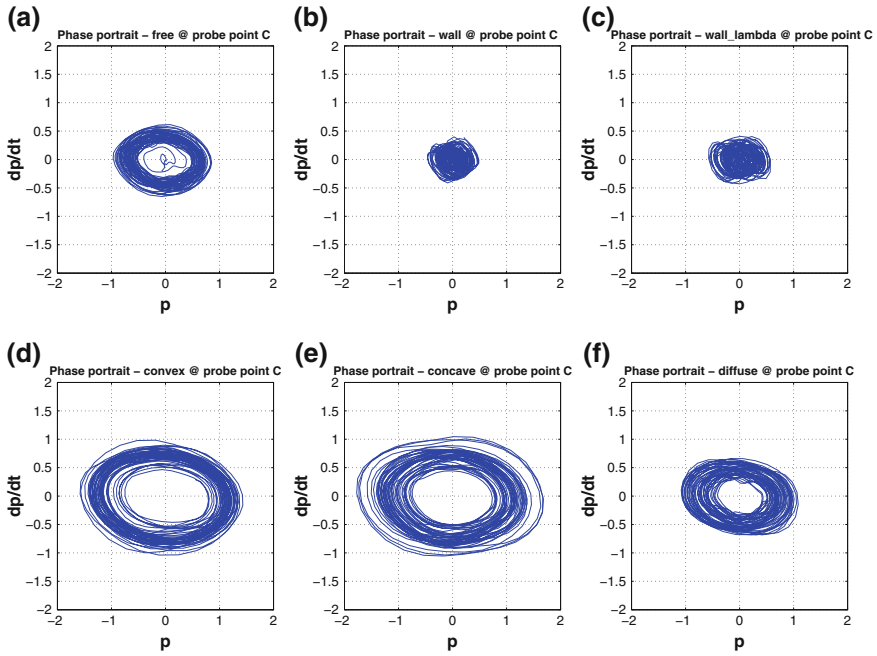


Fig. 16 Phase portraits of the scenarios **a** *free*, **b** *wall*, **c** *wall_lambda*, **d** *convex*, **e** *concave* and **f** *diffuse*. From the full simulation timespan of $t_{sim} = 100$ ms the part $t = 10-80$ ms is depicted.

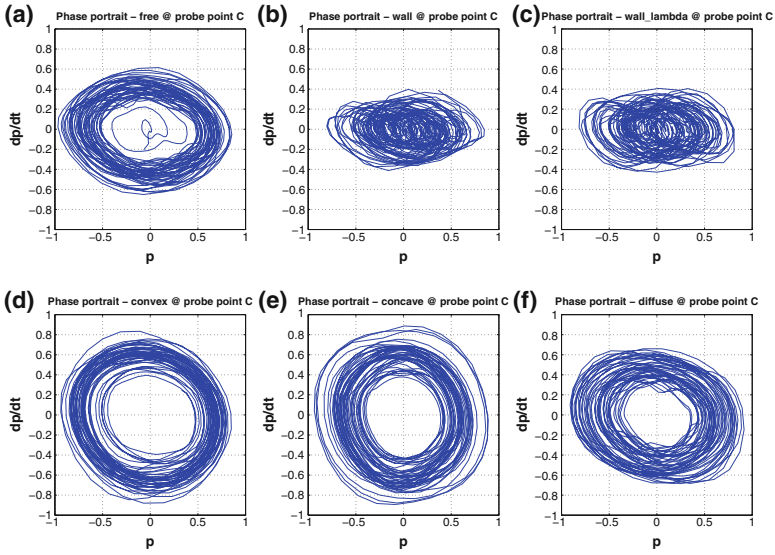


Fig. 17 Normalized phase portraits of the scenarios **a** *free*, **b** *wall*, **c** *wall_lambda*, **d** *convex*, **e** *concave* and **f** *diffuse*. The timespan depicted is t = 10−80 ms

oscillator systems. Nevertheless, it is recognizable that all scenarios represent self-sustained oscillator systems.

The phase portraits of scenarios *wall* and *wall_lambda* show strong disturbances, which prevent the formation of larger amplitudes. The doubling of period mentioned above can be clearly seen in Figs. 17b, c.

The trajectories in the phase portrait of scenario *convex* are elliptical, and circular in the normalized representation in Fig. 17d, thus resembling that of a harmonic oscillator. The course of the trajectories of scenario *concave* rather reminds one more of a tilted square than of a circle, this being especially the case with the outer trajectories.

Perhaps this is the matter of a second oscillatory regime with larger amplitudes into which the system transitions in case sufficient energy is available. This might be caused by the focusing of the incoming sound waves, an issue that, however, will not be treated within the scope of this article. In any case, a substantial change within the self-excited oscillatory behavior can be observed in comparison to the reference scenario *free*.

The spatial orientation of the trajectories in the phase portrait of scenario *diffuse* deviates considerably from those of the other scenarios. The phase portrait appears to be tilted clockwise. The pressure's zero crossings do not coincide with the points of maximum pressure change, implying a phase shift between these values in the order of $\Delta\phi \approx \pi/8$.

The phase portraits attest to the considerable influence of different spatial geometries on the sound generation process within the organ pipe. The nonlinearities within the system are significantly affected by the surrounding spatial geometry. This can be observed as deformations of the trajectories' regimes, as shown in the phase portraits. The non-linear oscillations within the organ pipe are being downright 'reshaped' through the influence of the different geometry. Depending on the type of spatial geometry, certain frequencies are either being accentuated or attenuated, leading to a significant change in the organ pipe's timbre.

4 The Feedback Effect of Swell Chamber Geometries

In this section, the feedback effect of the swell chamber geometry on the sound generation and sound radiation of an organ pipe is addressed. This part of the study evaluates the results of two numerical simulations, taking into account different swell chamber boundary conditions. In the first part, the geometries applied are briefly discussed. Following this, the computational execution of the simulation is outlined. In the third part, the results of the simulation are analyzed and compared to the results of the reference scenario *free*. This survey concludes with a brief summary as well as an outlook towards future work.

The numerical simulation is performed on the basis of a wooden stopped organ pipe having a quadratic cross-section and a fundamental frequency of $f = 700\,\mathrm{Hz}$,

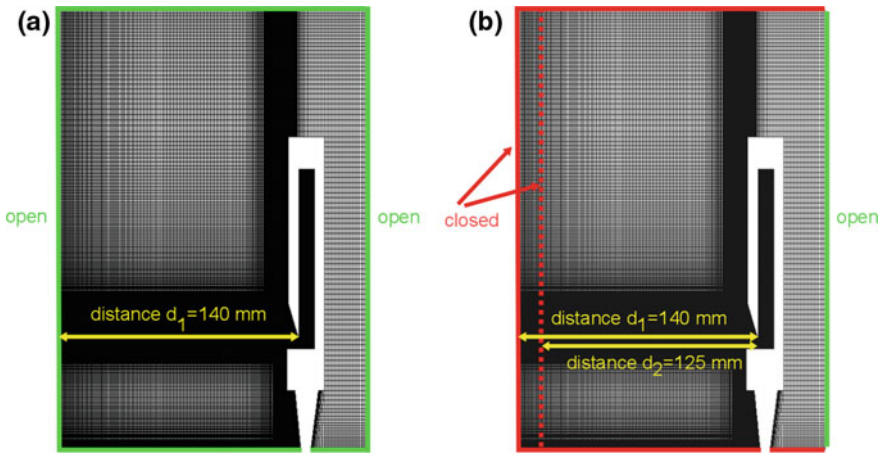


Fig. 18 Schematic description of the boundary conditions of the numerical simulations, **a** reference scenario *free*, **b** the scenarios *swell_140* mm and *swell_125* mm, with a wall distance of 140 mm and 125 mm to the pipes mouth

as shown in Fig. 1a. The organ pipe was custom-built by organ makers Schuke Orgelbau Potsdam GmbH [2] and provided for measurement use. The geometry of the organ pipe as well as the near-field ambient space are translated into a pseudo-3D computational grid by the procedure described in Sect. 3; the computational grid is shown in Fig. 1b.

The key data of the computational grid used are those of scenario *free* (cf. Table 1). The thermo-physical properties are summarized in Table 2.

In terms of boundary conditions for the room, two different configurations are implemented. Their geometries are equivalent to those of scenarios *wall* and *wall_lambda*, in this case, however, exhibiting three acoustically inert surfaces: the floor, the wall opposite the cut-up, and the ceiling. Only the numerical boundary behind the organ pipe remains open. The boundary conditions of the reference scenario *free* and the configurations *swell_140* mm and *swell_125* mm are shown in Fig. 18a, b. No-slip boundary conditions are chosen for the organ pipe's inside and outside walls as well as for the walls of the swell chamber. For the organ pipe's windway, i.e., the inlet and the open room boundaries herein referred to as outlet, boundary conditions are chosen that ensure conservation of mass.

4.1 Numerical Simulations of an Organ Pipe within Swell Chambers

Each run of the simulation generates a data volume of approximately 80 GB. From the total amount of generated data, animated sequences are compiled for the

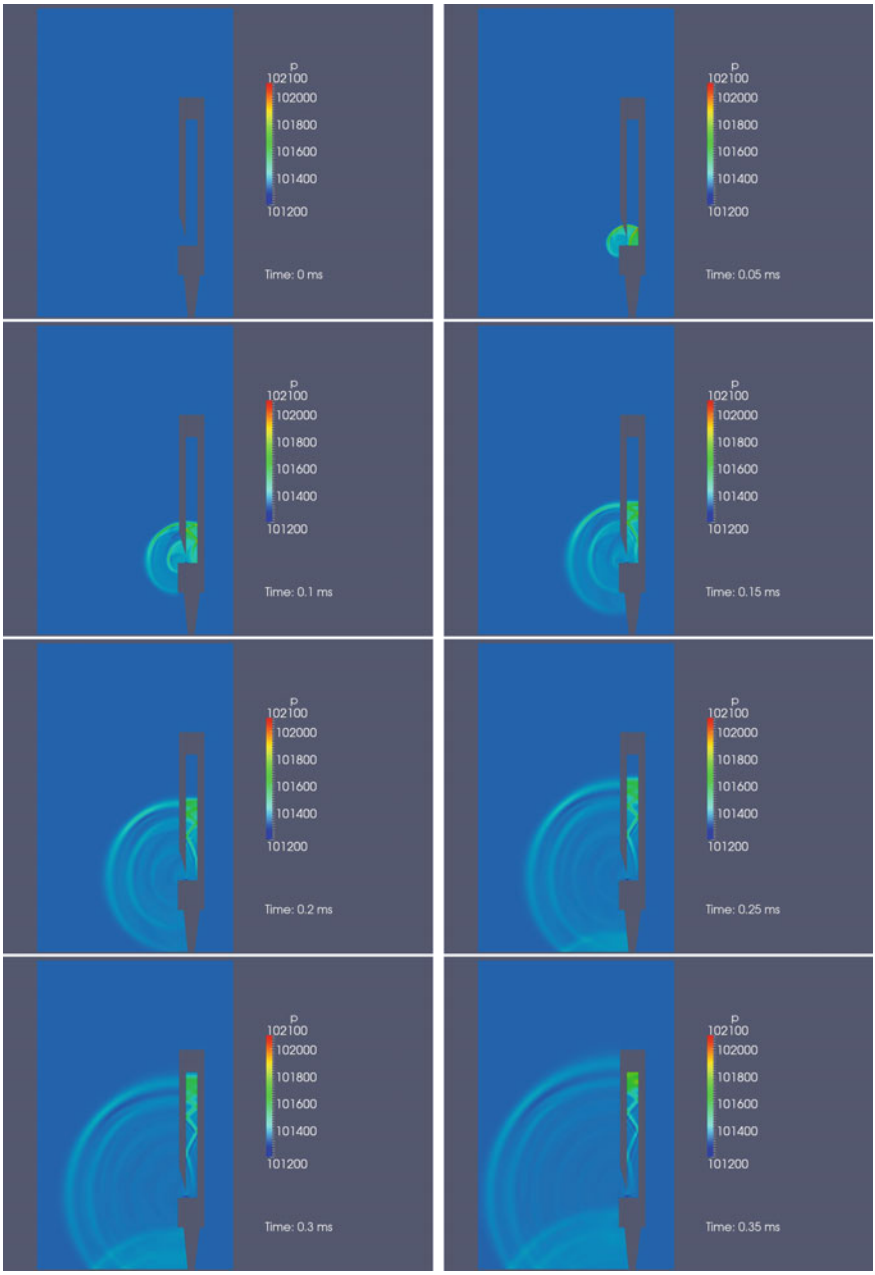


Fig. 19 Sequence $t = 0-0.35$ ms of the numerical simulation of scenario *swell_125* mm. Shown is the pressure p. Depicted is the initial excitation process of the organ pipe and the radiation of a sound wave into the swell chambers space

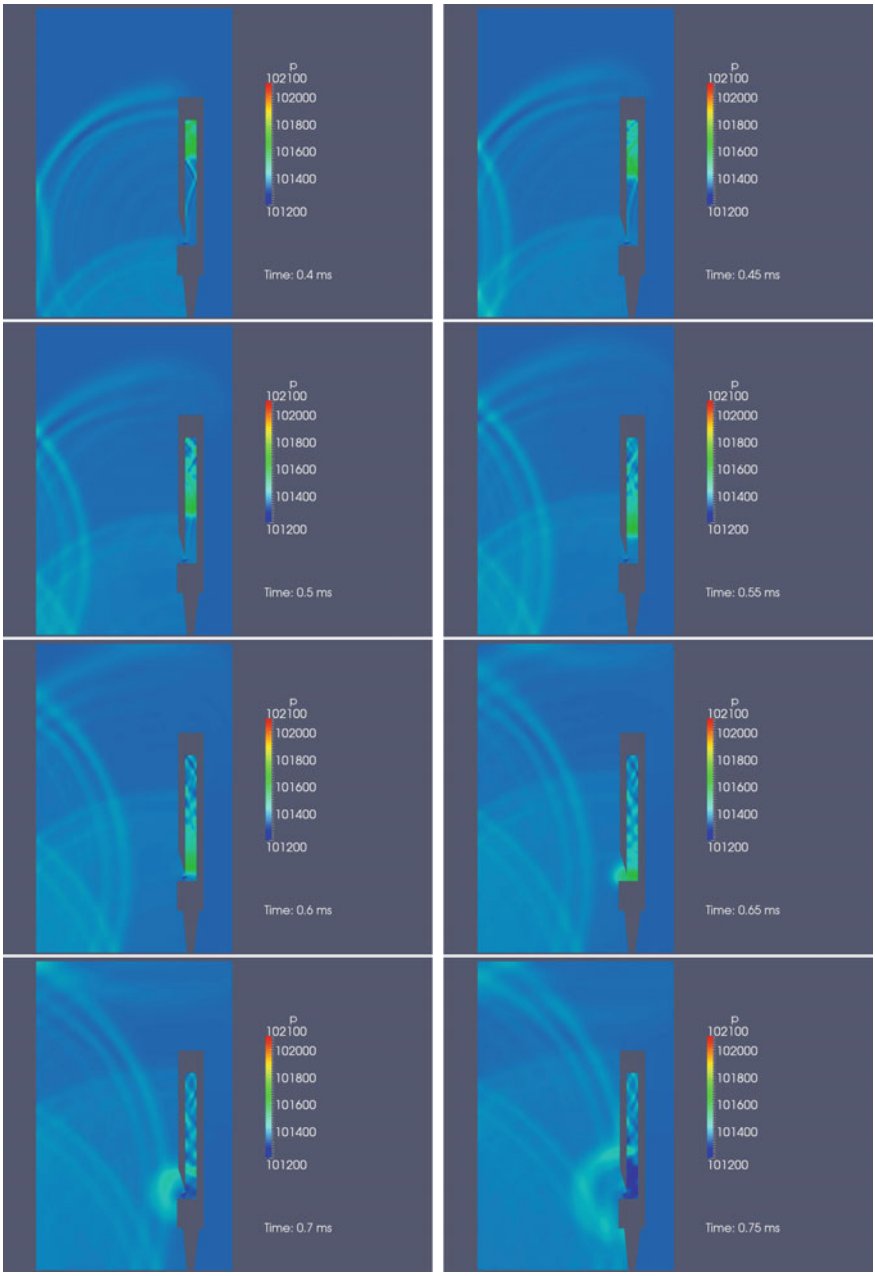


Fig. 20 Sequence $t = 0.4 - 0.75$ ms of the numerical simulation of scenario *swell_125* mm. Shown is the pressure p. Depicted is the initial excitation process of the organ pipe, the radiation of a sound wave into the swell chamber's space, reflection at the swell chamber's walls and back propagation

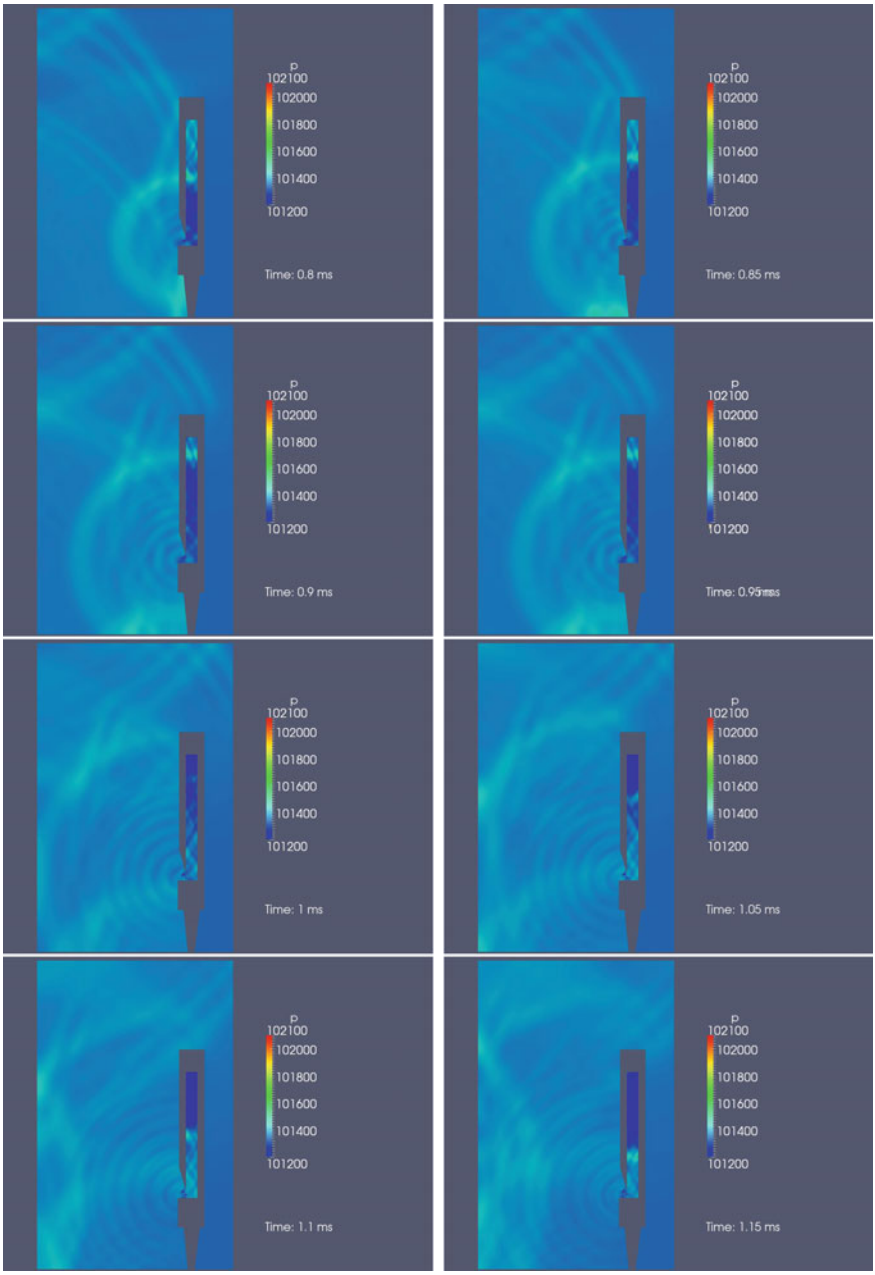


Fig. 21 Sequence $t = 0.8 - 1.15$ ms of the numerical simulation of scenario *swell_125* mm. Shown is the pressure p. Depicted is the initial excitation process of the organ pipe, the radiation of a sound wave into the swell chamber's cavity, reflection at the swell chamber's walls and back propagation

physical values, pressure p, velocity magnitude $|U|$, and turbulent kinetic energy k, to allow for a detailed qualitative evaluation of the processes within the organ pipe as well as in its surrounding environment.

Figures 19, 20 and 21 show sequences of the computed numerical simulation for the swell chamber configuration *swell_125* mm. Depicted is the initial excitation process. Color-coded is the pressure p in the range of 101200–102100 Pa.

4.2 Analysis

For the purpose of analyzing the simulation process, sound pressure level spectra are generated at probe point *C*, i.e., in the upper part of the sealed resonator, as well as at cross-section *cs*0, representing the cut-up. Probe point *C* is selected for its close proximity to the pressure maximum at the upper end of the sealed resonator. The data for cross-section *cs*0 (272 sample points per physical value) are coarse grained by spatial averaging and dividing by the number of sample points.

Thus a SPL-spectrum of the organ pipe's total sound radiation is obtained, reduced to a single point. Both data volumes are of particular interest with regard to an evaluation of the feedback effect of the surrounding room on the sound generation and sound radiation of the organ pipe, for they allow for the investigation of, on the one hand, the effects deep within the organ pipe, in the resonator, and, on the other hand, of the organ pipe's radiation characteristics. Both locations are either very difficult to reach for the purpose of experiment, or not at all. Figure 22 shows

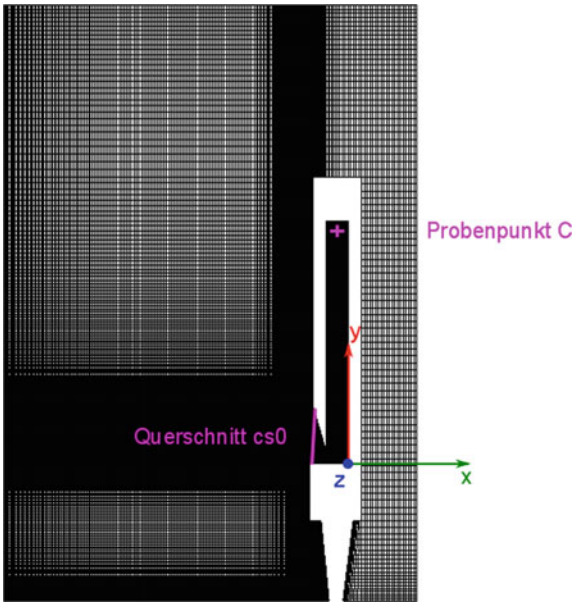


Fig. 22 Position of the probe point *C* and the cross-section *cs*0 within the mesh

the location of cross-section *cs*0 and the position of probe point *C* (Position [$x = -4.5$ mm, $y = 96$ mm]) within the computational grid.

4.3 Sound Pressure Level Spectra Inside the Organ Pipe

Analyzed are the sound pressure level spectra sampled at the probe point *C* for the three numerical simulation runs *free*, *swell_140* mm and *swell_125* mm, shown in Fig. 23a. One observe the fundamental as well as the higher harmonics up to the 13th.

The first harmonics up to the 5th are depicted in Fig. 23b–d.

With simulation *free*, the fundamental oscillation's frequencies as well as the higher harmonics deviate less than 5 % from the values established by experiment.

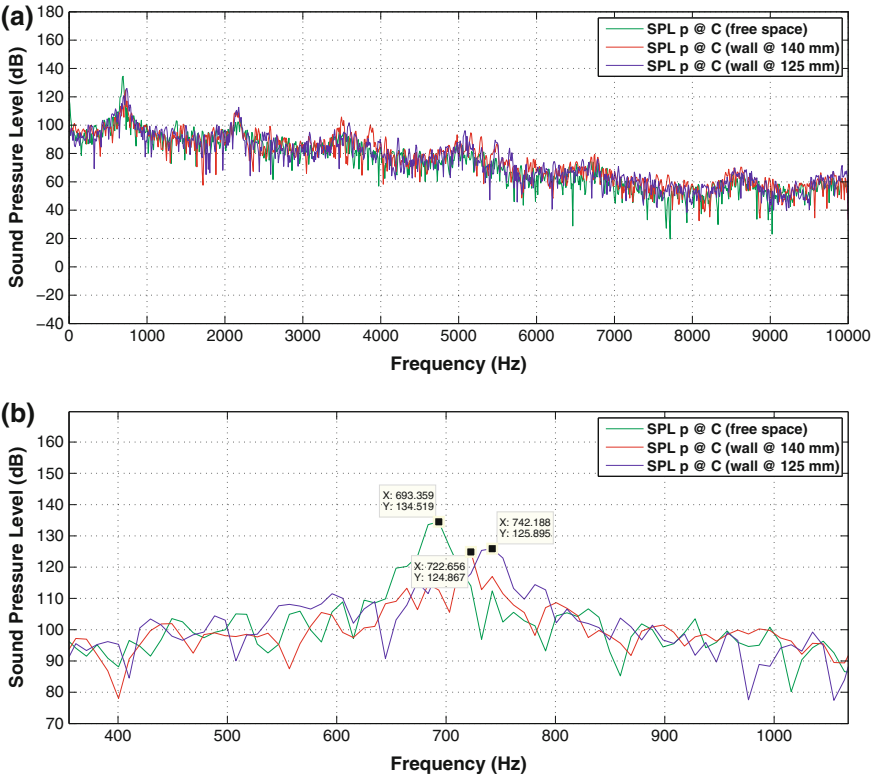


Fig. 23 SPL-spectra sampled at probe point *C* inside the resonator. **a** SPL-spectra of *free*, *swell_140* mm and *swell_125* mm. **b** Detailed view on the SPL-spectra in the range of the fundamentals. **c** Detailed view on the SPL-spectra in the range of the 3rd harmonics. **d** Detailed view on the SPL-spectra in the range of the 5th harmonics

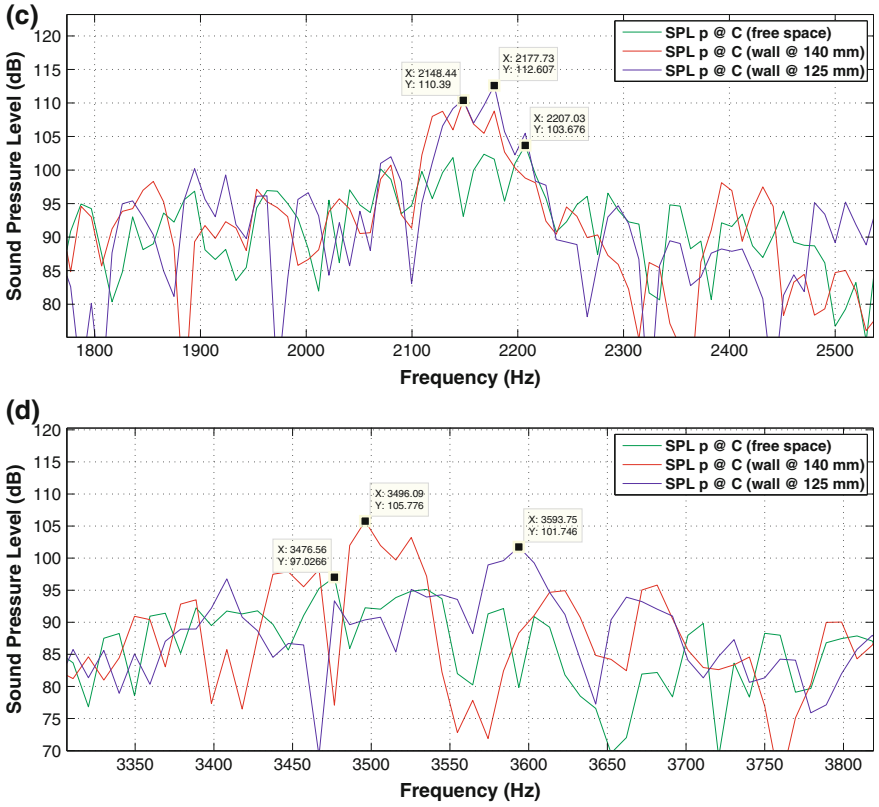


Fig. 23 (continued)

The simulation thus depicts the behavior of the organ pipe with great precision. The high degree of accuracy remains unaffected when boundary conditions are altered by the introduction of reflective surfaces, as in simulations *swell_140* mm and *swell_125* mm. This is another important quality characteristic of the presented simulations and attests of the robustness and reliability of the techniques applied.

On close examination of the area of fundamental oscillations in Fig. 23b, a frequency shift relative to the reference of approx. 29 ± 5 Hz and 49 ± 5 Hz, respectively, can be detected in simulation runs *swell_140* mm and *swell_125* mm. Furthermore, the reflecting walls reduce the sound pressure level of the fundamental oscillations by approximately 9.9 and 8.6 dB, respectively. This means that the reflecting walls of the swell chamber significantly affect both frequency and amplitude of the organ pipe. Within a swell chamber that is closed on three sides, the organ pipe undergoes an increase in frequency of the fundamental oscillation as well as a decrease in loudness.

4.4 Higher Harmonics

For the perceived sound quality of an organ pipe, the response to higher harmonics is of profound importance. Consequently, 3rd and 5th harmonics will now be evaluated, with simulation run *free* again serving as reference case. The sound pressure level of the 3rd harmonic of *swell_140* mm is 6.7 dB higher than in *free*. The sound pressure level of simulation *swell_125* mm is even higher at 8.9 dB. This means that reflecting surfaces lying in the direction of sound radiation lead to an energy transfer from the fundamental to the 3rd harmonic. This transfer even increases as the distance to the opposing wall is decreased, as is the case with configuration *swell_125* mm. This effect is observable also in the case of the 5th harmonic. The increase is, however, smaller in *swell_125* mm than in *swell_140* mm. In comparison to simulation *free*, the differences amount to 8.7 dB for *swell_140* mm and 4.7 dB for *swell_125* mm. The transfer of energy towards higher harmonics is a highly non-linear process, hitherto not well understood. It is part of current research. To make statements with an even higher degree of precision, the simulated duration would have to be at least doubled, which is, in principle, technically feasible.

4.5 Spatially Averaged Sound Pressure Level Spectra of the Cut-up Region

Next is the analysis of the spatially averaged sound pressure level spectra at cross-section *cs*0, spanning the organ pipe's cut-up. By spatially averaging and dividing by the number of sample points (272) the pressure data, the organ pipes total radiation characteristic, reduced into a single point, is obtained.

The sound pressure level spectra thus obtained may be interpreted as the radiation characteristics of the organ pipe, seen as an acoustical point source. The sound pressure level spectra across cross-section *cs*0 are shown in Fig. 24a–c.

In comparison to the sound pressure level spectra of probe point *C*, the frequencies of the fundamental oscillation as well as those of the higher harmonics emerge even more clearly. The increase in frequency of the fundamental, caused by the feedback effect of the swell chamber's reflecting walls, is analogous in behavior to the observations on probe point *C* within the resonator.

The averaged sound pressure levels are slightly lower than within the resonator. This is explained by friction losses at the boundary layer of the inner walls of the resonator affecting the propagating sound wave. At 115.6 dB, the sound pressure level of simulation run *free* corresponds to the experimentally determined values for this organ pipe. The sound pressure levels of the fundamental oscillations in simulation runs *swell_140* mm and *swell_125* mm are, at 103.8 dB and 103.4 dB, respectively, reduced by 11.7 dB and 12.2 dB compared to *free*, equivalent to a decrease by a factor of 4.

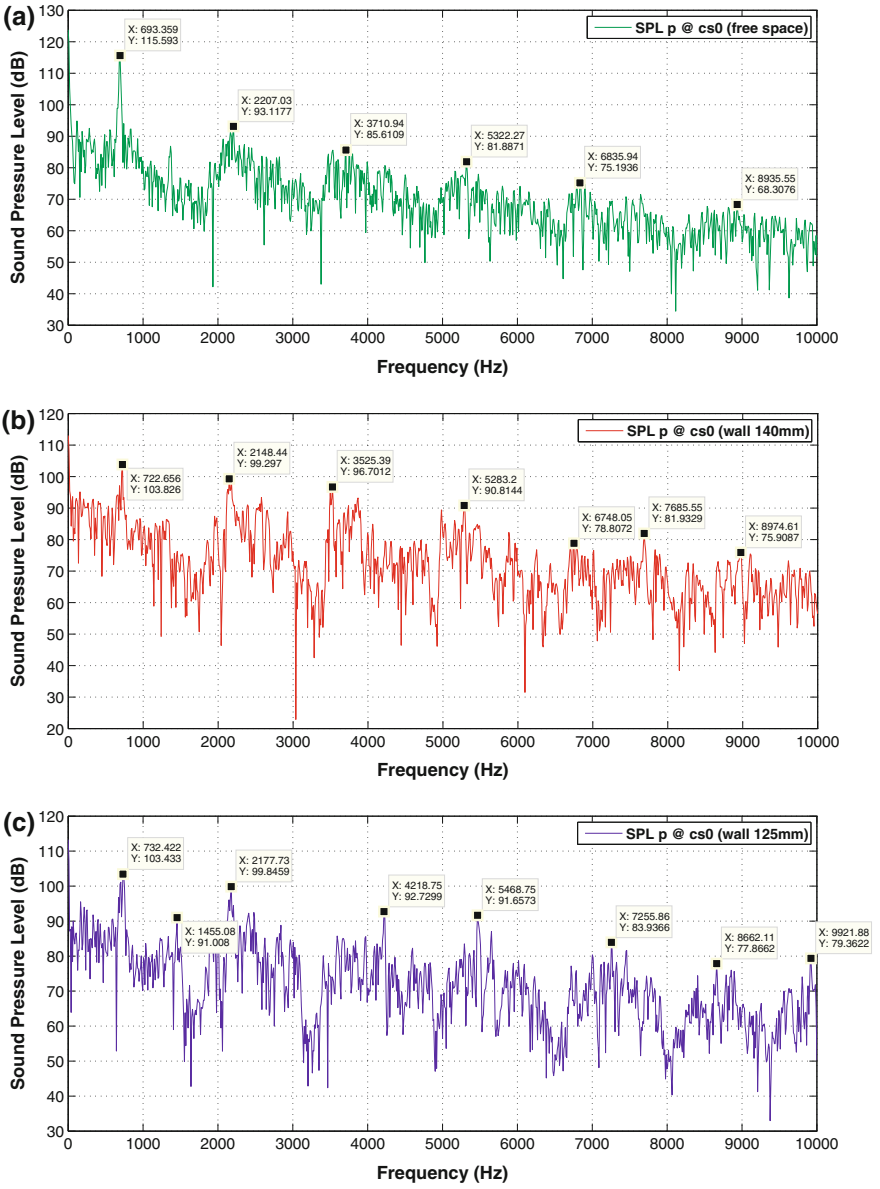


Fig. 24 Spatially averaged sound pressure level spectra sampled at cross-section *cs*0 of **a** *free*, **b** *swell_140* mm, **c** *swell_125* mm

The energy transfer from the fundamental towards the higher harmonics is more clearly visible as with probe point *C*. It raises the 3rd harmonic by 6.2 dB in *swell_140* mm and by 6.7 dB in *swell_125* mm. In case of the 5th harmonic, an

increase of 11.1 dB and 7.1 dB, respectively, and of 8.9 dB and 9.8 dB, respectively, in case of the 7th harmonic, are observed.

Remarkable, again, is the appearance of a 2nd harmonic in the sound pressure level spectra, as averaged across the cut-up.

In summary it can be stated that the fundamental oscillation is significantly affected by the swell chamber. Depending on the proximity of the organ pipe to an opposing acoustically inert wall, a frequency shift towards higher frequency occurs. Hereby, the fundamental oscillation experiences massive damping, while the higher harmonics are being enhanced. With regard to a more in-depth investigation of these phenomena, further experimentation as well as additional numerical simulations are encouraged.

4.6 *Auto-synchronization of the Organ Pipe by Swell Chambers Feedback*

Indications of auto-synchronization of the organ pipe by means of the wall's feedback effects are to be found when analyzing the spatially averaged velocity components across cross-section *cs*0. In Fig. 25a–c, velocity components v_y and v_x are shown, running nearly tangentially and nearly transverse across cross-section *cs*0. It can be seen that, with regard to sound radiation, the fluid dynamical processes and the acoustical processes of sound generation can be well separated. Here, only the acoustical phenomena are discussed. It can clearly be seen that v_x contains mainly the particle velocity, being the velocity component running transverse across cross-section *cs*0.

In the reference scenario, *free*, particle velocity proceeds mostly undisturbed for 80 ms. The simulations incorporating the swell chamber, *swell_140* mm and *swell_125* mm, on the other hand, exhibit interference in the form of period-doubling and beating. These disturbances are caused by sound signals propagating back from the wall with a certain distance-dependent phase shift, interfering with the oscillating jet. Additionally, it has to be taken into account that the particle velocity experiences a sudden phase shift of $\phi_{reflect} = \pi$ at being reflected at the wall, yet sound pressure does not! This is of importance as it is the particle velocity that acts upon the jet.

From theoretical considerations [8, 9] as well as experiments [7, 10], it is known that organ pipes may synchronize under certain circumstances. The interdependencies of sound field and flow field of the jet are discussed in detail in [6].

Mutual cancellation of the particle velocities occurs in the case of both the particle velocity generated on the resonator side, acting upon the jet, and the particle velocity of the reflection, impinging from the outside upon the jet, being offset by π. As a result, the deflections of the jet come to a standstill, a phenomenon known as oscillation death, or quenching [8].

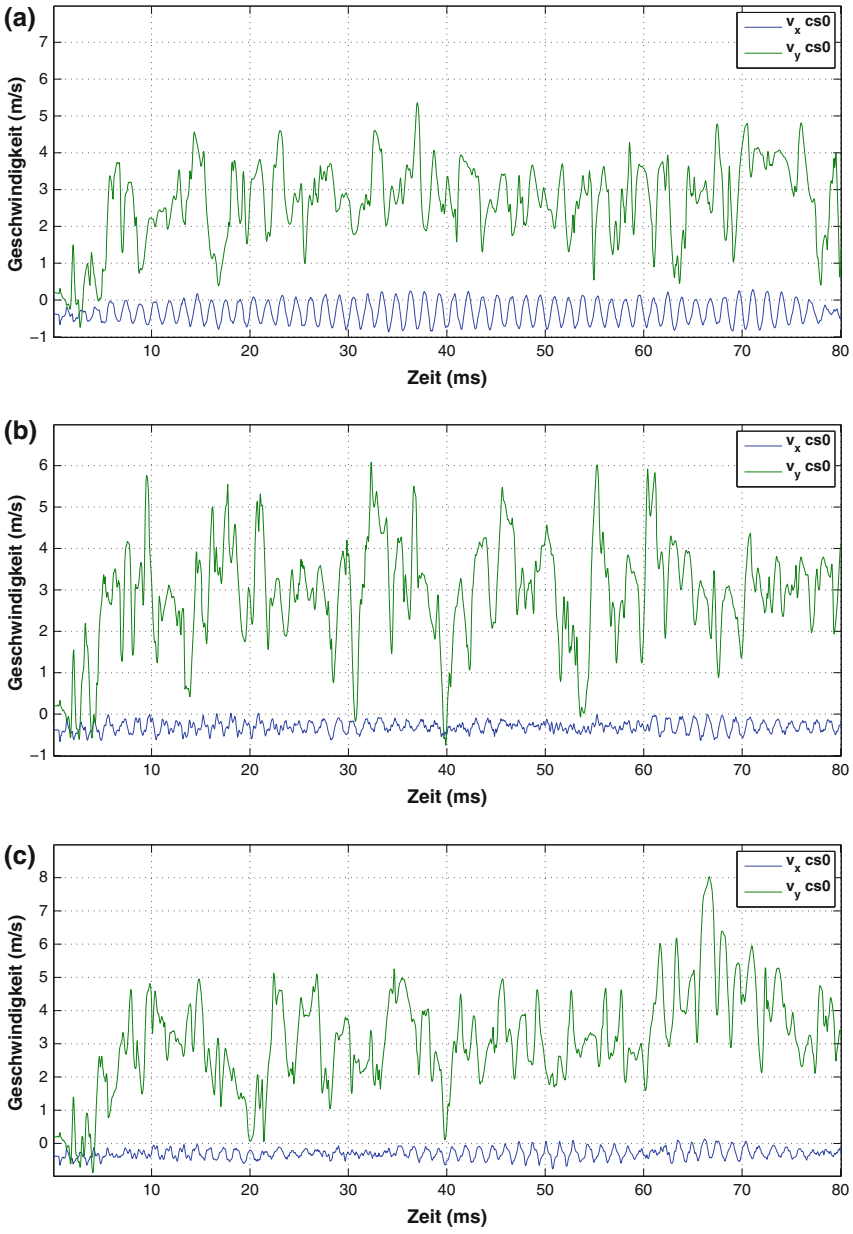


Fig. 25 Spatially averaged velocity components v_x and v_y at cross-section *cs*0 for **a** the reference scenario*free*, **b** the scenario *swell_140* mm and **c** the scenario *swell_125 mm*. Note that the component v_x is orthogonal to the main time-averaged flow direction of the jet. It can be seen clearly that v_x, compared with v_y, is much more periodic and with ca. 1/10 lower amplitudes than v_y. This indicates that v_x is the carrier of fraction of the sound of the whole signal while the v_y component mainly represents the irregular fraction of the signal, namely the flow of the jet

Based on the results of the analysis, the hypothesis is stated that auto-synchronization, i.e., a synchronization of the jet with its own radiated, reflected, and time delayed sound signal of the same frequency, can occur if the phase difference $\Delta\phi$ of the radiated as well as of the reflected sound signal approaches $\Delta\phi = 0$ and $\Delta\phi = \pi$, respectively. This is equal to the propagation lengths of the sound signal in the surrounding room of multiples of λ and $\lambda/2$, respectively. This hypothesis is supported by the decrease in sound pressure level for a distance of propagation of approximately $\lambda/2$ (*swell_125* mm), as discussed above.

5 Summary

This chapter dealt with the influence of complex spatial geometries on the sound generation and sound radiation of an organ pipe. The diverging responses of the different wall geometries with regard to the initial sound pressure wave could be clearly distinguished. The processes of sound generation and sound radiation of the organ pipe are significantly affected by the respective spatial geometry of the circumambient room.

Affected are the sound pressure levels of the fundamental oscillation and of the higher harmonics, as well as their frequencies. Particularly noticeable seems to be the variance in the distribution of the sound pressure levels across the fundamental and the higher harmonics. Depending on the spatial geometry, the levels of different frequencies are either accentuated or attenuated, leading to various distinct acoustic patterns for one and the same organ pipe. By selecting a specific spatial geometry, the organ pipe's acoustic pattern may be deliberately altered.

The results of contemplating the effects of swell chambers on sound generation and sound radiation with regard to organ pipes are summarized in note form:

- The acoustic pattern of the organ pipe is significantly affected by the swell chamber.
- The frequency shift caused by the swell chamber's geometry depends on the distance between the organ pipe and the opposite wall.
- Here, a significant distance-dependent decrease in loudness of the fundamental oscillation occurs, as well as an accentuation of higher harmonics.
- The sound pressure level spectra show an energy transfer towards higher harmonics.
- The organ pipe's ability to auto-synchronize with its own radiated sound signal is feasible.

Furthermore, it has been shown that the effects of the circumambient spatial geometry on the organ's sound may be analyzed and displayed by means of numerical simulation in conjunction with the analysis software developed for this purpose, prior to actual organ design and construction. The numerical simulations executed herein allow for further analyses of, e.g., defined cross-sections within the

computational grid as well as of additional physical values that decisively affect the mechanisms of sound generation within the organ pipe and its interdependencies with external parameters.

Acknowledgments The author wants to express his gratitude for inspiring discussions with A. Pikovsky, M. Rosenblum, M. Abel and R. Gerhard from University of Potsdam. Many thanks to Alexander Schuke Orgelbau Potsdam GmbH for their active help in pipe and wind supply construction. J. L. Fischer was supported by ZIM, grant "Synchronization in Organ Pipes".

Bibliography

1. Adelung, W.: Einführung in den Orgelbau. Breitkopf & Hartel, Leipzig (1991)
2. Alexander Schuke Orgelbau Potsdam GmbH. URL http://www.schuke.com/ (2016)
3. OpenFOAM®—The Open Source Computational Fluid Dynamics (CFD) Toolbox Organization—OpenCFD Limited. URL http://www.openfoam.com/ (2016)
4. Morse, P.M., Ingard, K.U.: Theoretical Acoustics. Princeton University Press, Princeton, NJ (1968)
5. Schlichting, H., Gersten, K.: Boundary-layer theory. Springer, Berlin (2003)
6. Fischer, J. L.: Nichtlineare Kopplungsmechanismen akustischer Oszillatoren am Beispiel der Synchronisation von Orgelpfeifen, Ph.D. thesis, available at University of Potsdam (2014)
7. Fischer, J. L.: Über Synchronisationsphänomene nichtlinearer akustischer Oszillatoren, Diploma-thesis, available at University of Potsdam (2012)
8. Pikovsky, A., Rosenblum, M., Kurths, J.: Synchronization—A Universal Concept in Nonlinear Science. Springer, Berlin (2001)
9. Rayleigh, J.W.S.B.: *The* Theory of Sound, Republished 1945 by Dover Publications, New York (1896)
10. Abel, M., Ahnert, K., Bergweiler, B.: Synchronization of sound sources. Phys. Rev. Lett. **1030**, 114301 (2009)

Author Biography

Jost Leonhardt Fischer has been a postdoctoral researcher in the Institute of Systematic Musicology at University of Hamburg, since 2014. His current research focuses on applications of nonlinear dynamics and oscillation theory in musical acoustics. Topics include, inter alia, synchronization phenomena in acoustical waveguides, nonlinearities in sound generation and sound radiation, investigations of the interplay of flows, turbulent layers and sound field as well as numerical simulations of the compressible Navier-Stokes equations. Jost Leonhardt Fischer studied physics at the University of Potsdam, Germany. In his Diploma thesis (2012) he investigated synchronization phenomena of nonlinear acoustic oscillators, from both a numerical and a theoretical perspective. In 2014 he received a Ph.D. in theoretical physics. In his Ph.D. thesis, he studied nonlinear coupling mechanisms of acoustic oscillators with a focus on synchronization of organ pipes.

Acoustical Modeling of Mutes for Brass Instruments

Shigeru Yoshikawa and Yu Nobara

Abstract Mutes for brass instruments are placed in the bell and change the instrument resonance characteristics. Since the pitch, loudness, and timbre are then affected, brass mutes are important for musical expression. Our main focus is on the acoustical modeling of the straight mute for the French horn and the cup mute for the trumpet. The validity of our numerical analysis is confirmed by the results on the hand stopping and the stopping mute for the horn. An application of our modeling method to other trumpet mutes is furthermore demonstrated.

1 Introduction

There are various mutes for many musical instruments. These mutes are well designed to fulfil their role in playing the corresponding musical instruments. Let us see the mutes for the violin and piano at first. A violin mute is usually three-pronged, and clipped to the bridge in order to weaken and darken the sound. It adds an extra mass to the bridge and shifts the resonances to lower frequency. Also, this added mass strengthens the impedance mismatch between strings and bridge. As a result, energy transfer to the soundboard is reduced, particularly in higher frequencies. Similar mutes are made for viola, cello, and double bass. The violin mute appeared in the 17th century and Henry Purcell (1659–1695) used it for his works [1].

S. Yoshikawa (✉)
1-27-22 Aoyama, Dazaifu 818-0121, Japan
e-mail: shig@lib.bbiq.jp

Y. Nobara
Fujitsu Ten Ltd, 3-7-2-304 Mikagehommachi, Higashinada, Kobe 658-0046, Japan
e-mail: yu.nobara@outlook.com

A. Schneider (ed.), *Studies in Musical Acoustics and Psychoacoustics*,
Current Research in Systematic Musicology 4,
DOI 10.1007/978-3-319-47292-8_5

Modern grand pianos have no muting device exactly, however, they have soft pedal, called '*una corda*'. This pedal shifts the action sideway so that the hammer misses hitting one of two or three strings per note. Using this soft pedal in *fortessimo*, the player may produce peculiar dark and shady sounds. Like this, mutes largely extend player's expression capability.

On the other hand, mutes for brass instruments are nearly conical and put in the bell. Since the mute almost closes the bell, acoustical radiation of lower frequencies is reduced. However, higher frequencies are emphasized by resonances within the mute that is open at the narrow end and is closed at the wide end. As a result, blurry and thin tones are produced. An ancient form of trumpet mute is considered to have been used probably before the violin mute was known [1]. Different types of mutes appeared in the 19th century and created unique tonal characteristics required by composers.

There are some acoustical or musicological references on brass mutes. Ancell [2] investigated acoustical effects of the cornet mutes. Backus [3] suggested electrical analog circuits of the mutes for the trumpet, and further discussed the acoustics of hand stopping and stopping mute for the horn. Smith [4] extended the brass mute research from the acoustical and musicological viewpoint, and gave effective suggestions for the future research. Watts [5] tried to compare acoustical differences between open (no hand-in-bell), hand-in-bell, and hand-stopping effects in the horn. Natalie [6] carried out the experimental research on the player's right hand (using three replica hands) and several horn mutes, and further developed psycho-acoustical research on the auditory feedback provided to horn players.

Acoustical and musical effects produced by several mutes for brass instruments will be discussed in more detail in this chapter. A chief objective is to present adequate acoustical models of mutes, which have not been advanced yet expect for electrical analog models of Backus [3], for the French horn and trumpet.

2 Hand-in-Bell and Hand-Stopping in the French Horn

2.1 Effects of Hand in Horn Bell

The French horn is different from other brass instruments in the use of player's right hand. In normal playing of the horn the player places his or her right hand in a position within the bell (the flared end of the instrument) to produce a desired tone with correct intonation. This normal situation is called *hand-in-bell* in this chapter. The hand is slightly cupped, but the bell end is mostly open to secure sufficient radiation of sound. The readers can see typical hand position and shape for playing the French horn in Refs. [5–8].

The acoustical difference between open bell (no hand-in-bell) and hand-in-bell is clearly shown in the temporal waveform and the corresponding frequency spectrum of the radiated sound. Such examples are given in Fig. 1a, b. The fourth resonance

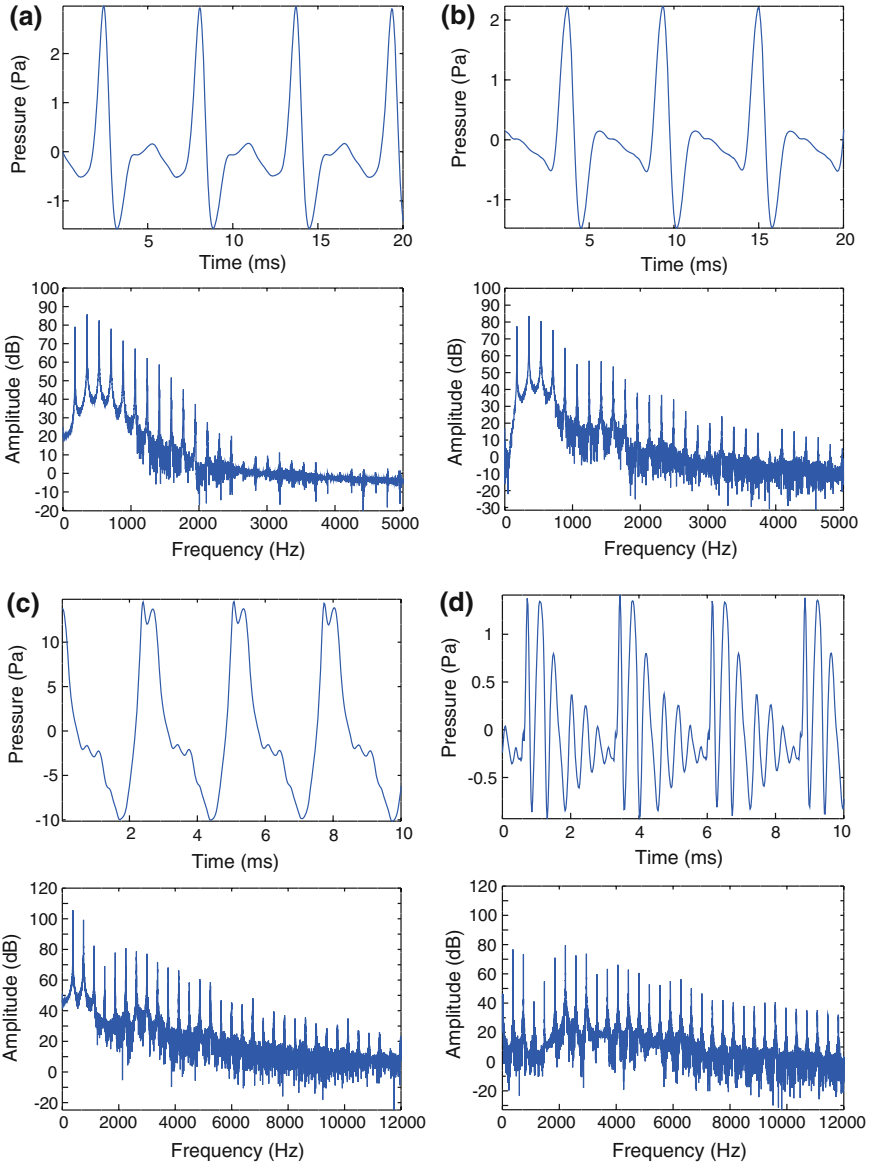


Fig. 1 Sound examples in French horn (in the key of F). **a** F_3 (178 Hz) by open horn; **b** F_3 (179 Hz) by normal (hand-in-bell) horn; **c** $F^{\#}_4$ (376 Hz) by hand-stopped horn; **d** $F^{\#}_4$ (370 Hz) by muted horn with a stopping mute. In each case the temporal waveform (*the upper frame*) and its frequency spectrum (*the lower frame*) are illustrated

mode (tone F_3) of a natural horn (in the key of F) manufactured by the Lawson Brass Instruments Inc., in the US was played by the same player with almost the same loudness (in *mezzoforte*). Note that the pressure and dB values are calibrated. The open bell indicates the cutoff frequency above which the standing wave for the resonance is very weakly formed along the entire instrument and the radiation efficiency (the ratio of the radiated pressure to the internal pressure) tends to be 1.0. This cutoff frequency for brass instruments $f_c = \omega_c/2\pi$ is approximately given by [9]

$$f_c \approx c/\pi a, \tag{1}$$

where c denotes the speed of sound and a the bell radius at the end. This f_c nearly equals 720 Hz when c = 344 m/s (22 °C) and $2a$ = 30.5 cm are applied. The radiation efficiency almost increases as 6 dB per octave below f_c [9]. However, it is difficult to confirm this f_c from the tonal spectrum which continues to about f = 2500 Hz in Fig. 1a. This might be due to high radiation efficiency above f_c. If the impedance curve that defines the resonance characteristics of the horn is observed, it may be inferred that f_c is located around 700 Hz (see Figs. 5a and 11a).

On the other hand, the player's hand in the bell restricts the bell area and causes substantial decrease in a. Hence, acoustic radiation of high-frequency harmonics is reduced and inversely acoustic reflection of those harmonics near the bell end is increased. The hand-in-bell situation can thus form stronger high-frequency standing waves between the mouthpiece and bell as shown in Fig. 1b. As a result, the horn player can accurately select the individual resonance mode even exceeding the 16th mode above the fundamental, whereas the trumpet and trombone play up to about the 8th mode (the horn uses harmonic series up to an octave higher) [10].

2.2 Hand Stopping and Stopping Mute

When the player's hand is placed deeper into the bell and the bell exit is almost completely closed, the pitch, loudness, and timbre of the radiated sound produced by strongly blowing the horn are drastically changed from those in normal playing. This playing technique is referred to as *hand stopping* (*gestopft* in German) [7, 8]. The timbre of the stopped tone is characterized as *metallic brittle and rough* [11] in contrast to *brassy* timbre of the tones in normal playing. An example of this stopped tone is shown in Fig. 1c. The pitch is $F^{\#}{}_4$ and the loudness is largely increased in comparison with Fig. 1b in normal playing. It is distinctive that the stopped tone produces strong higher harmonics even above 10 kHz.

Concerning the stopped-tone frequencies of the F horn, Backus [3] explained that all the resonance frequencies except the first and second moved down, and each of those resonances (order n) ended up a semitone above the frequency where the neighboring *lower* resonance (order $n - 1$) had been originally located. In other words, the third resonance falls near the original second in normal playing, the

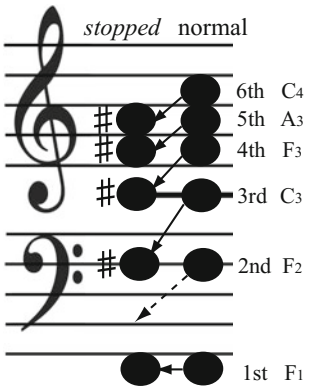


Fig. 2 Pitch change by hand stopping [19]. The *right column* shows the 1st to 6th mode of the tone in normal (hand-in-bell) playing; the *left column* the corresponding mode of the *stopped* tone

fourth near the original third, and so on. However, the resulting frequencies of all resonances move up a semitone compared with those of resonances in original normal playing (see Fig. 2). This moving down of the resonance mode is due to the interstices formed around the hand (or the fingers thrust hard into the bell) that terminate the horn with a bigger inertance. On the other hand, as Backus [3, 10] points out, the reason for a semitone rise in pitch is often misunderstood as a result that the air column has been shortened by filling its part with the hand as in Morley-Pegge [12].

In spite of his informative study on the input impedance in hand stopping, Backus [3] did not fully explain the physical cause of the stopped-tone timbre, but just suggested that the channels between the fingers acted as high-pass filters and increased the amplitudes of higher harmonics relative to those of lower harmonics. Actually, possible physical causes of the metallic timbre of the stopped tone seem to be various. First, the steepening of the waveform (or the formation of the shock wave) due to nonlinear propagation through a straight tube in the trumpet and trombone [13–15] comes to the mind. Also, the influence of wall vibration on brass sound should be considered [16–18]. The energy transfer to the bell from the player's lip oscillation [16] or from the internal modal pressure in the resonant air column [18] may be the dominant source to generate the wall vibration in playing brass instruments. Moore et al. [17] experimentally examined the wall-vibration effect by immersing the bell and horn sections in sand. The physical cause of the stopped-tone timbre will be discussed in the following sections in more detail.

The horn player uses a stopping mute (see Fig. 3) that has the effect similar to hand stopping. The bell is sealed off by the cork portion of the mute, but the acoustic pressure inside the bell can be radiated through a narrow throat. Its tonal characteristics are shown in Fig. 1d. The waveform of the muted horn is very different from that of the hand-stopped horn. Although the sound pressure is 1/10 of the hand-stopped horn, higher harmonics are very plenty. Generally, the stopping mute raises the tonal pitch a semitone, and the player needs to lower the pitch with transposition. Then, such stopping mutes are called *transposing* mutes. At present, *non-transposing* stopping mutes (without pitch change) have been developed.

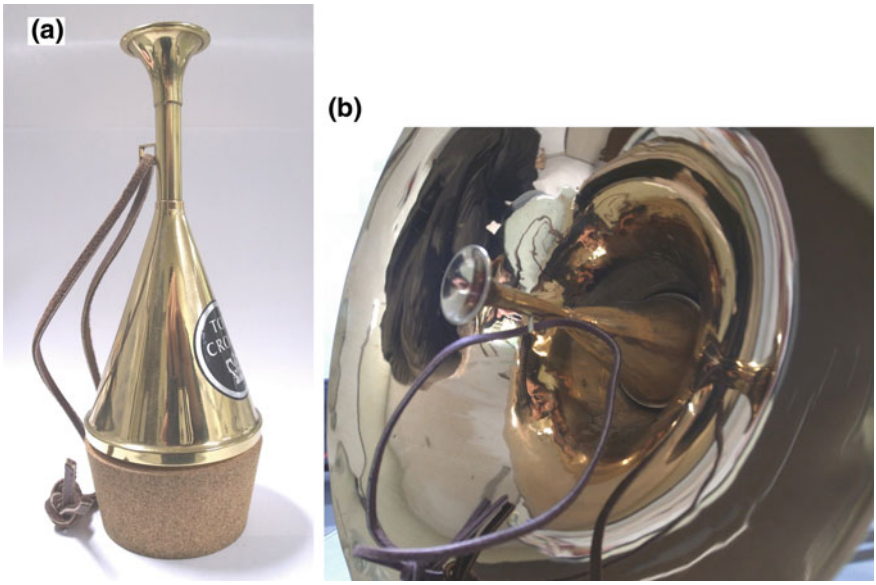


Fig. 3 A stopping mute (Tom Crown make) for the French horn. **a** Total view of the mute; **b** external view when the mute is inserted in the horn

Acoustical modelling of a transposing stopping mute as well as hand stopping will be carried out in the next section before considering acoustical models of various mutes for brass instruments.

2.3 Acoustical Modeling of Hand-in-Bell and Hand Stopping

As mentioned above, the horn player's right hand can be used in three ways: Not placed in the bell [the horn in this case is called an *open* (*no hand-in-bell*) horn]; placed in the bell as usual [called a *normal* (*hand-in-bell*) horn]; inserted into the bell almost completely (called a *stopped* horn). These three types of hand placements are simply modeled as in Fig. 4a–c respectively [19].

The cross section of the Lawson's F natural horn is shown in Fig. 4a. This is an open horn example. The total length is 385 cm:145 cm of the flaring bell section; 182 cm of the cylindrical section; 51.6 cm of the lead pipe section; 6.4 cm of the mouthpiece. Also, the diameter of the bell is 30.5 cm at the end, and that of the cylindrical section is 1.2 cm.

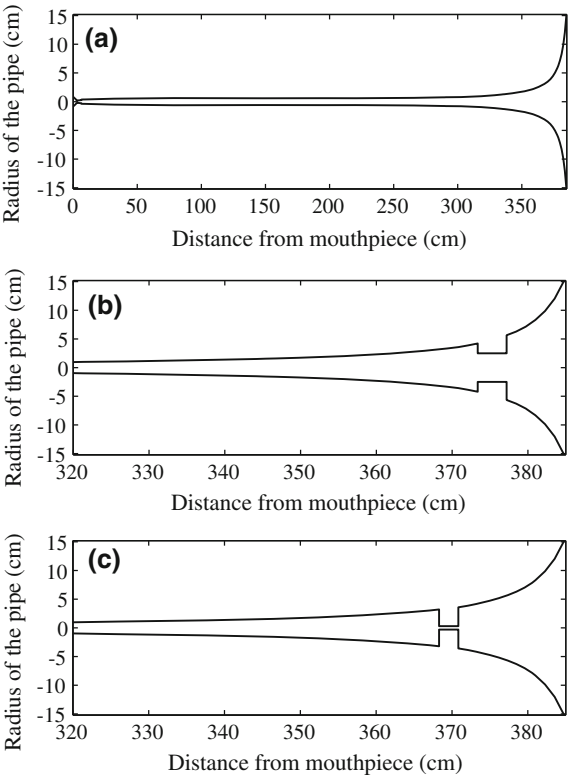


Fig. 4 Simple modeling of the player's right hand [19]. **a** The cross section of the natural horn (open horn) used to calculate the input impedance; **b** The enlarged view of the cross section of the normal-horn model with the right hand in bell; **c** The enlarged view of the cross section of the hand-stopped-horn model

Because our objective is to execute straightforward calculation of the normal and stopped horns based on simple models, the real shape of the right hand is replaced with a thin plate with an orifice. The area of the orifice corresponds to that of the interstice between the bell wall and the player's hand. For our calculation, it is supposed that the distance from the bell to the plate is 7.6 cm, the plate diameter is 8.4 cm (equal to the internal diameter of the bell there), the plate thickness is 3.8 cm, and the orifice diameter is 5.0 cm. These dimensions are defined so that the input impedance yielded by the calculation is consistent with measurement results by Dell et al. [6, 20]. The enlarged view of the cross section of our normal-horn model is illustrated in Fig. 4b.

Our stopped-horn model shown in Fig. 4c has a plate with a smaller orifice, which is set deeper in the bell. The distance from the bell to the plate is 15.2 cm, the plate diameter 6.4 cm, thickness 2.5 cm, and orifice diameter 6 mm. The orifice diameter is nearly equal to that of a narrow cylindrical pipe involved in a stopping mute (cf. Fig. 10).

2.4 Input Impedances of the Open, Normal, and Stopped Horns

Following the method by Caussé [21], the bore is first divided into a series of small sections (cylinders and truncated cones) to apply simple acoustical theory. In the nth section from the bell, input pressure $p_{n,\mathrm{in}}$, input volume velocity $q_{n,\mathrm{in}}$, output pressure $p_{n,\mathrm{out}}$, and output volume velocity $q_{n,\ \mathrm{out}}$, are related by the transmission matrix (or the T-matrix) T_n as follows:

$$\begin{pmatrix} p_{n,in} \\ \mathrm{q}_{n,in} \end{pmatrix} = T_n \begin{pmatrix} p_{n,out} \\ \mathrm{q}_{n,out} \end{pmatrix} = \begin{pmatrix} A_n & B_n \\ C_n & D_n \end{pmatrix} \begin{pmatrix} p_{n,out} \\ \mathrm{q}_{n,out} \end{pmatrix} \tag{2}$$

Each element of T_n is given by equation developed by Mapes-Riordan [22] by including the effects of visco-thermal losses (see Table II in [22]). In our actual calculation we used cylindrical and conical elements in steps of 5–10 mm. From the continuity of p and q at boundaries between adjacent elements, the following relation is finally obtained:

$$\begin{pmatrix} p_M \\ \mathrm{q}_M \end{pmatrix} = \prod_n T_n \begin{pmatrix} p_B \\ q_B \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} p_B \\ q_B \end{pmatrix} \tag{3}$$

where p_{M} and q_{M} are the pressure and the volume velocity at the input end of the mouthpiece, and p_{B} and q_{B} are those at the output end of the bell. Thus, the input impedance $Z_{\mathrm{in}} = p_{\mathrm{M}}/q_{\mathrm{M}}$ of the horn is expressed as

$$Z_{in} = \frac{Ap_B + Bq_B}{Cp_B + Dq_B} = \frac{AZ_L + B}{CZ_L + D} \tag{4}$$

where $Z_{\mathrm{L}} = p_{\mathrm{B}}/q_{\mathrm{B}}$ corresponds to the radiation (or load) impedance at the bell end, which may be approximated by a piston set in an infinite baffle since the presence of the baffle seems to have a relatively small effect [9].

Also, the transmission function giving the ratio of the output sound pressure p_{out} (or the bell pressure p_{B}) to the input sound pressure p_{in} (or the mouthpiece pressure p_{M}) is expressed as

$$H(f) = 20\log|p_{out}/p_{in}| = \frac{1}{AD - BC}\left(D - \frac{B}{Z_L}\right) = D - \frac{B}{Z_L} \tag{5}$$

where it should be noted that $AD - BC = 1$ is generally satisfied for the transmission matrix.

The absolute magnitude of Z_{in} of Eq. (4) is calculated for the open horn (Fig. 4a) and the normal-horn model (Fig. 4b), respectively. The calculated results are compared in Fig. 5a using the blue line (the open horn) and the red line (the normal-horn model) [19]. The impedance curves show almost no change below the

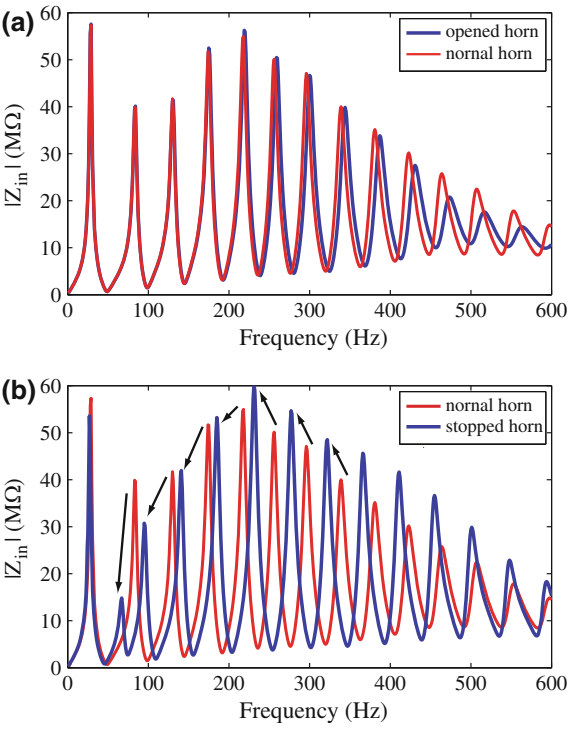


Fig. 5 Calculation of input impedances of the horns [19]. **a** Comparison of normal (hand-in-bell) horn with open (no hand-in-bell) horn; **b** comparison of stopped (hand-stopped) horn with normal horn

fourth mode. However, the hand placed in the bell yields increased maxima as well as reduced minima in the higher frequency range (above the ninth mode). This means a significant increase in *playability* of the upper modes due to the reflection at the hand in the bell. Also, the frequency of the upper modes is a little lowered due to the mass effect of the hand, but there is no appreciable change in the lower modes. The calculated result of Fig. 5a agrees with experimental data taken by Backus with the plasticine replica hand (see Fig. 10 in [3]) and by Watts [5] with a model hand. The blue line also suggests the existence of the cutoff frequency f_c of Eq. (1) around 700 Hz.

Another comparison of the input impedance is given in Fig. 5b between the normal-horn model (red line) and the stopped-horn model (blue line) [19]. It is distinct that the hand stopping markedly diminishes the second harmonic as Backus [3] found using his rubber stopper. As indicated by the arrows, a downward shift of all peaks due to the hand stopping is observed. As a result, the frequency ratio R of the $(n + 1)$st mode of the stopped-horn model to the nth mode of the open horn is around 100 cents (a semitone) except for $n = 2$ as summarized in Table 1.

It should be noted that the transmission function [defined as $H = 20\log|p_B/p_M|$ in Eq. (5)] of the stopped-horn model does not indicate a specific high-pass filter characteristic expect for a resonance peak around 6.8 kHz (see Fig. 14 in Sect. 3.3 or Fig. 3a in Ref. [19]). This peak corresponds to the resonance frequency of an open pipe of 2.5 cm, i.e., the orifice length of the modeled plate.

Table 1 Resonance frequencies f_r given by the input impedances of the open horn, the normal-horn model, and the stopped-horn model

		Open	Normal		Stopped horn			Frequency ratio R	
Mode	Tone	f_r (Hz)	f_r (Hz)		Mode	Tone	f_r (Hz)	Mode	R (cent)
I	(Pd.)	29	29	→	I′	(Pd.)	27		
II	F_2	84	85	→	II′	(almost disappeared)			
III	C_3	131	131	→	III′	$F^{\#}{}_2$	95	III′/II	213
IV	F_3	175	175	→	IV′	$C^{\#}{}_3$	141	IV′/III	127
V	A_3	219	221	→	V′	$F^{\#}{}_3$	185	V′/IV	96
VI	C_4	259	262	→	VI′	$A^{\#}{}_3$	231	VI′/V	92
VII	$D^{\#}{}_4$	300	304	→	VII′	$C^{\#}{}_4$	277	VII′/VI	116
VIII	F_4	344	348	→	VIII′	E_4	322	VIII′/VII	123
IX	G_4	387	392	→	IX′	$F^{\#}{}_4$	366	IX′/VIII	107
X	A_4	431	438	→	X′	$G^{\#}{}_4$	411	X′/IX	104

The frequency ratio R between the resonance frequencies at the closest corresponding modes of the open horn and the stopped-horn model is also given

2.5 Pressure Distribution Along the Horn

The pressure distribution along the horn is calculated to obtain a qualitative understanding of the pitch change caused by hand stopping. Although Backus has described that the pitch change by hand stopping is due to the inductive termination at the bell [3], a more satisfactory explanation seems to be possible. When Eq. (3) is applied, the internal sound pressure p_k and the volume velocity q_k on the output end of the kth short pipe counted from the bell (k = 1, 2, 3, …) are given as

$$\begin{pmatrix} p_k \\ q_k \end{pmatrix} = \prod_{n}^{k-1} T_n \begin{pmatrix} p_B \\ q_B \end{pmatrix} = \begin{pmatrix} A_k & B_k \\ C_k & D_k \end{pmatrix} \begin{pmatrix} p_B \\ q_B \end{pmatrix}. \tag{6}$$

Thus, the internal sound pressure distribution is calculated as [19]

$$p_k = p_B \left(A_k + \frac{B_k}{Z_L} \right), \tag{7}$$

where a radiated sound pressure p_B is arbitrarily given to know a relative distribution pattern. The calculation result of the pressure distributions of F_3 (the 4th mode of the open horn, supposing p_B = 5 Pa) and $C^{\#}{}_3$ (the 3rd mode of the stopped-horn model, originally F_3, supposing p_B = 1 Pa) is illustrated in Fig. 6 [19].

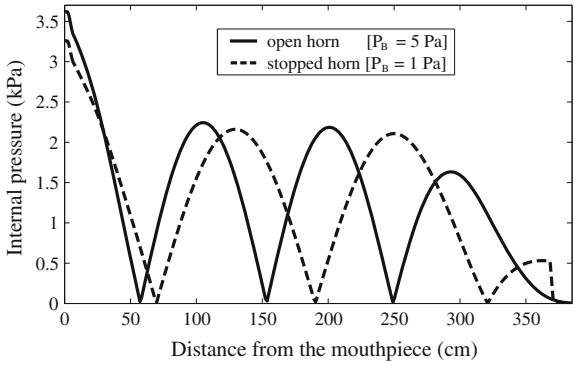


Fig. 6 Calculated internal pressure distribution along the horn [19]. The 4th mode of the open horn, F_3 under p_B = 5 Pa at a distance of 385 cm from the mouthpiece (*solid line*) and the 3rd mode of the stopped-horn model, $C^{\#}{}_3$ (originally F_3) under p_B = 1 Pa (*dashed line*). The internal pressure in hand stopping at 370 cm from the mouthpiece is 12 times larger than that in the open horn at the same position

It is evident that the pressure mode pattern along the horn in hand stopping changes from a "closed-open" pipe pattern to a "closed-closed" pipe pattern. In other words, both pressures at the mouthpiece end and at the bell end are the maxima in hand stopping. This effect means a larger inertance at the bell; this makes the wavelength in the horn longer than the original one and shifts the corresponding original mode (order n) in the "closed-open" pipe down to the next lower mode (order $n - 1$) in the "closed-closed" pipe. That is the essential cause of the puzzling pitch *descent* (shown by the arrow in Fig. 2) by hand stopping. However, if the comparison is made at the same mode order n, hand stopping produces pitch *ascent* of nearly a semitone as shown in Fig. 2 and in Table 1 (remember that the second mode of the stopped horn is almost disappeared). This is because the wavelength of the nth mode in the "closed-closed" pipe is a little shorter than that of the nth mode in the "closed-open" pipe.

2.6 Physical Cause of Metallic Timbre by Hand Stopping

It is evident that the waveform of the radiated stopped tone ($C^{\#}{}_4$) shown in Fig. 4b of Ref. [19] indicates *rapidly corrugating change* (minutely indented waveform), although such a change is not so appreciable in Fig. 1c (such a change is seen in Figs. 8 and 9). On the other hand, the normal tone shown in Fig. 1b has no such a change. Furthermore, the spectrums of the stopped tones given in Ref. [19] and Fig. 1c indicate the same characteristics as have been denoted by Meyer [11], that is, an amplitude reduction of the harmonics from 1 to 2 kHz and an emphasis of higher harmonics held up to 10 kHz. This strongly suggests that the penetrating

metallic timbre of the stopped tone can be caused by the rapidly corrugating change observed in the waveform. However, this rapidly corrugating change is also found in the waveforms of wall-vibration velocity in hand stopping (not shown here, see Fig. 4 in Ref. [19]). Therefore, we have to examine two possibilities which cause the rapidly corrugating change: (1) direct radiation from the bell wall and (2) other mechanisms such as nonlinear wave steepening along the bore [13–15] and player's lip vibration [23].

Let us directly confirm whether the wall vibration contributes the timbre of the stopped tones or not by strongly damping the horn body. If the metallic stopped tone is generated by the wall radiation, the tonal metallicness should be removed when the horn bell and the pipes are completely damped [24]. As shown in Fig. 7, the bell of the natural horn was mounted in a wooden enclosure, and the pipe of the horn was placed in a box. If these two boxes were filled with a quantity of sand, wall vibrations of the horn should be strongly damped. The natural horn was played in hand stopping in an anechoic room of Graduate School of Design, Kyushu University, Japan. The radiated sound pressure and the wall vibration at the bell edge were measured under four conditions [19]: No sand in the boxes (called the *free* condition), with the sand poured into the bell section only (the *bell-damped* condition), with the sand into the pipe section only (the *pipe-damped* condition) and with the sand into both sections (the *fully damped* condition). The player was asked to keep his right hand in the same position and to play at the same volume in each condition.

The amplitude of the vibration velocity at the bell in the fully damped condition is reduced to about 1/20 of that in the free condition, and the amplitude of the spectral envelope of the velocity is also reduced by 20–40 dB in the frequency range up to 10 kHz due to the damping [19]. Therefore, if the wall vibration

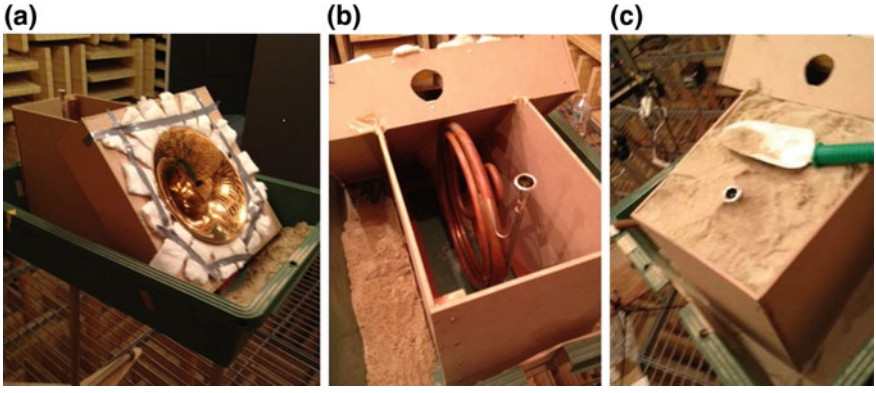


Fig. 7 The procedure of immobilization experiment of the horn body. **a** The bell mounted in a wooden enclosure using vibration absorber; **b** the pipe section placed in a box; **c** the box filled with sand

radiates the metallic stopped tone, some clear differences in stopped tones should be observed between damping conditions.

The radiated sound pressures of the stopped tones in four damping conditions are illustrated in Fig. 8. All the measured sound waveforms still show characteristic minute wave corrugation, even though horn body is strongly damped by the sand. Also, their spectral envelopes (see Fig. 11 in [19]) do not indicate definite differences regardless of whether the horn is damped or not. Although the prime characteristic of the metallic stopped tone is the peak around 3 kHz as suggested by Meyer [11], the immobilization of the horn body cannot remove that peak. Hence, it may be considered that the wall vibration does not primarily affect the stopped sound.

It should be thus examined whether nonlinear propagation along the bore generates the rapidly corrugating waveform in hand stopping. In the context of nonlinear propagation or wave steepening, the corrugating waveform (or change) may be adequately replaced with the *wave corrugation*. The Burgers equation predicts that the shock wave is generated if the length of the cylindrical pipe is longer than the critical distance [13, 14, 25]:

$$x_c = \frac{2\gamma P_{at} c}{(\gamma+1)[dp_M/dt]_{max}}, \tag{8}$$

where γ = 1.4 is the Poisson ratio, P_{at} the mean atmospheric pressure, c the sound speed, and p_M the mouthpiece pressure. Even if the length of cylindrical pipe of the instrument is shorter than x_c, the wave steepening is occurred and the high-frequency component of the radiated tone is increased [14, 25].

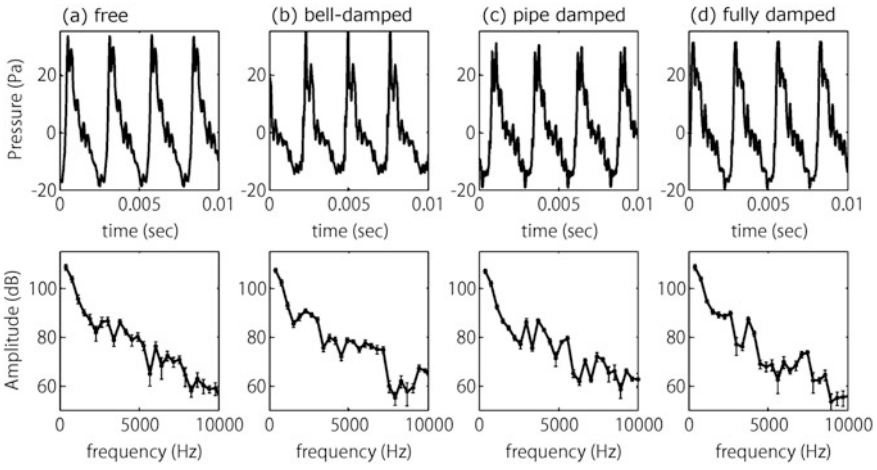


Fig. 8 Measured waveforms of the stopped tones ($F^{\#}{}_4$) under four damping conditions [19]. **a** Free; **b** bell-damped; **c** pipe-damped; **d** fully damped

The sound pressure p_M in hand stopping was measured by a pressure transducer (PCB 106B) that was attached to a mouthpiece (Yamaha 30D4). The signal from the transducer [powered by a conditioning amplifier (B&K 2693)] was amplified with a measuring amplifier (B&K 2636) and recorded on a computer with the sampling rate of 44.1 kHz. Also, the sound pressure radiated from the bell was measured using a microphone (B&K 4191). The measured waveforms of p_M and p_B in hand stopping are illustrated in Fig. 9a, c, respectively. The p_M in hand stopping is completely different from that in normal playing [23] and shows rough and arched peaks. The roughness of the peak seems to resemble that of p_M in almost fortissimo playing [25], and the arched peak is rather similar to the waveform measured in the trombone mouthpiece in loud playing [14].

The time derivative dp_M/dt in Eq. (8) is illustrated in Fig. 9b. This plot indicates that the maximum of dp_M/dt in hand stopping is much larger than that in normal playing. Equation (8) gives $x_c \approx 1$ m when $dp_M/dt = 40$ MPa/s, and the shock-wave formation is sufficiently possible in hand stopping. Furthermore, other smaller values of dp_M/dt at different peaks can cause the wave steepening. Therefore, it may be suggested that nonlinear propagation along the bore characterizes not only the *brassiness* of the fortissimo playing but also the *metallicness* of the stopped tone. Particularly, the wave corrugation characterizing the metallic stopped tones is possibly formed by a combination of many minute wave

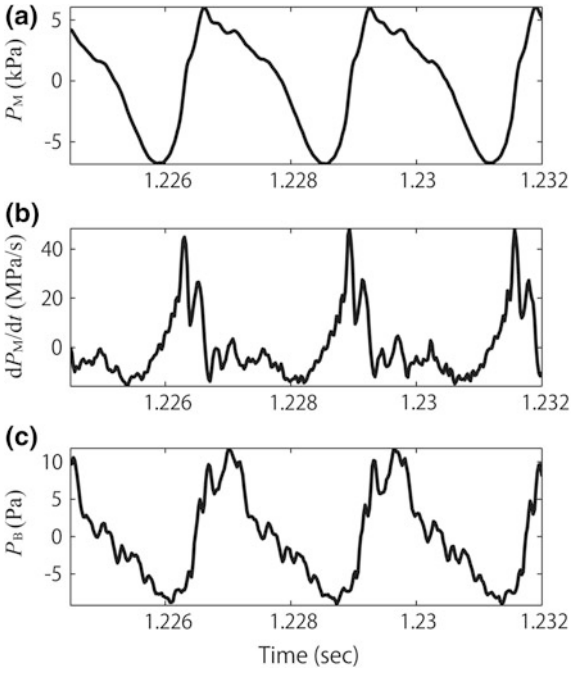


Fig. 9 Typical results of the mouthpiece-pressure measurement [19]. **a** Mouthpiece pressure in hand stopping; **b** rate of temporal change of the mouthpiece pressure; **c** radiated sound pressure of the stopped tone

steepenings as a result of nonlinear propagation. In order to confirm the responsibility of wave corrugations for the metallic timbre of the stopped tones, numerical simulations [26] of nonlinear propagation in time domain should be done in near future.

3 Stopping Mute for the French Horn

A stopping mute for the French horn was already shown in Fig. 3 and its tonal example was indicated in Fig. 1d. In this section, acoustical characteristics of the stopping mute are described in more detail [27].

3.1 Structure of Stopping Mute and Its Acoustical Characteristics

The cross section of the stopping mute (Tom Crown) is depicted in Fig. 10 by the solid line (also, see Fig. 12 in Ref. [3]) in comparison with our stopped-horn model drawn by the dashed line. The cork end of the mute comes in contact with the inner surface of the bell at a distance of about 366 cm from the mouthpiece end. The diameter of the mute cavity at this contact point is 5.15 cm, and its maximum diameter is about 6.5 cm. Total length of the mute is 18.5 cm. A narrow cylindrical pipe has its diameter of 0.8 cm and its length of 4.15 cm. Also, the diameter of the outer edge for sound radiation is 3.05 cm. This gives the radiation area equivalent to only 1 % compared with our open horn bell.

We can easily calculate the input impedance of the stopping-mute mounted natural horn on the basis of the transmission-line theory just as in the previous section. A comparison of $|Z_{in}|$ between the horn equipped with the stopping mute and the open (unstopped) horn is shown in Fig. 11a, and that between the horn

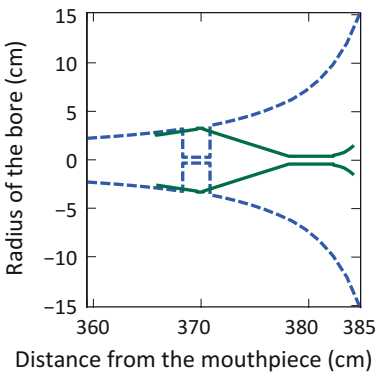


Fig. 10 Comparison of cross sections between the F natural horn equipped with the stopping mute (*solid line*) and the stopped-horn model (*dashed line*) given in Fig. 4c [27]

equipped with the stopping mute and the stopped-horn model in hand stopping is shown in Fig. 11b. Also, Table 2 indicates the peak frequencies of $|Z_{in}|$ for the horn equipped with the stopping mute and for the open horn.

The stopping mute brings about resonance frequencies similar to those given by the stopped-horn model. However, the frequency ratio R of the horn stopped by the stopping mute is a little lower than that of the stopped-horn model as indicated in Fig. 11b. This is also known from a comparison between Tables 1 and 2.

Backus [3] measured impedance curves of the horn with the stopping mute at varying degrees of insertion, and showed the lowering of the resonance frequencies and the diminishing of the second resonance in his Fig. 13. When the stopping mute was inserted with some leakage around it, resonances 2 and 3 have become a relatively low impedance double hump as shown in his Fig. 13d. Finally, when the stopping mute was completely inserted, the impedance curve around the original second resonance of the unstopped horn changed furthermore as shown in his Fig. 14. That is, a very small second peak appears to be generated and the third peak is apparent, although Backus [3] says that the second resonance has almost disappeared, being only a small hump on the lower frequency side of the third resonance. The third resonance is now about a semitone (100 cents) above the original second resonance of the unstopped horn. Likewise, the fourth, fifth, and higher resonances move to positions about a semitone above the frequencies of the original third, fourth, and higher resonances. This characteristic of the stopping

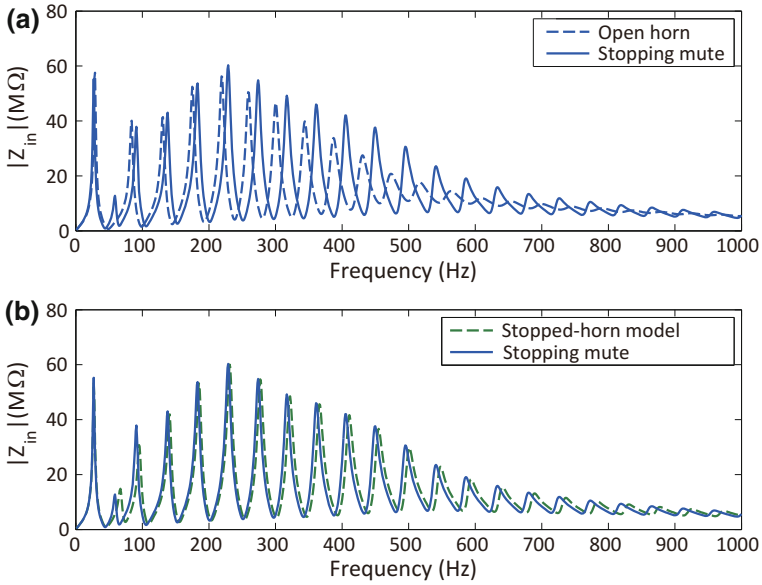


Fig. 11 Input impedance comparisons [27]. **a** Natural horn equipped with the stopping mute (*solid line*) versus the open horn (*dashed line*); **b** natural horn equipped with the stopping mute (*solid line*) versus the stopped horn modeling the hand stopping (*dashed line*)

Table 2 Resonance frequencies f_r of the input impedance of the open horn and the horn with the stopping mute

Open horn				Horn with the stopping mute			Frequency ratio R	
Mode	Tone	f_r (Hz)		Mode	Tone	f_r (Hz)	Mode	R (cent)
I	(Pd.)	29	→	I′	(Pd.)	27		
II	F_2	84	→	II′	$A^{\#}{}_1$	59		
III	C_3	131	→	III′	$F^{\#}{}_2$	91	III′/II	139
IV	F_3	175	→	IV′	$C^{\#}{}_3$	138	IV′/III	90
V	A_3	219	→	V′	$F^{\#}{}_3$	183	V′/IV	77
VI	C_4	259	→	VI′	$A^{\#}{}_3$	229	VI′/V	77
VII	$D^{\#}{}_4$	300	→	VII′	$C^{\#}{}_4$	274	VII′/VI	97
VIII	F_4	344	→	VIII′	E_4	317	VIII′/VII	95
IX	G_4	387	→	IX′	$F^{\#}{}_4$	361	IX′/VIII	83
X	A_4	431	→	X′	$G^{\#}{}_4$	405	X′/IX	79

The frequency ratio R between the resonance frequencies at the closest corresponding modes of these horns is also given

mute is quite the same as that of hand stopping if the small second peak of $|Z_{in}|$ in Fig. 11 may be interpreted as the almost disappeared second resonance in hand stopping (cf. Table 1). The calculated result given in Table 2 shows a quite good agreement with the measured result by Backus [3].

3.2 *Pressure Distribution Along the Horn*

The internal pressure distribution along the natural horn with the stopping mute and that without the stopping mute are illustrated in Fig. 12. The former corresponds to

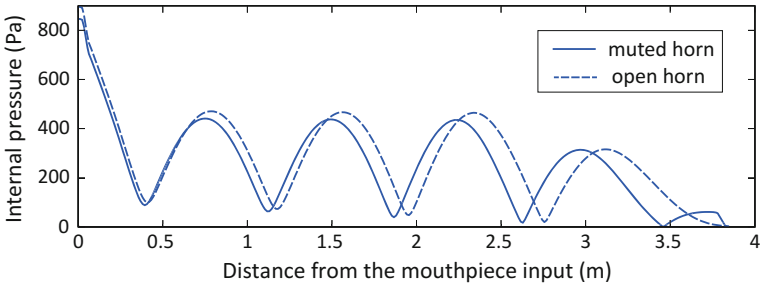


Fig. 12 Internal pressure distributions along the horn with the stopping mute (*solid line*, sixth mode, 229 Hz) and without the stopping mute (*dashed line*, fifth mode, 219 Hz) [27]. The radiated pressure p_B is assumed to be 1 Pa for the muted horn and 2 Pa for the unmuted (open) horn for better visibility

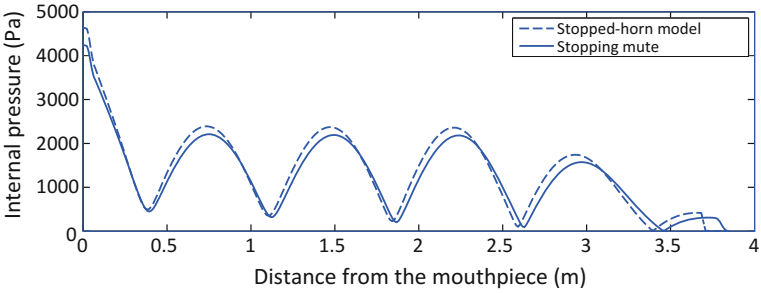


Fig. 13 Internal pressure distributions for the horn with the stopping mute (*solid line*, 229 Hz, p_B = 5 Pa) and for the hand-stopped-horn model (*dashed line*, 231 Hz, p_B = 1 Pa) concerning the sixth closed-closed pipe mode [27]

the sixth resonance (229 Hz) and the latter to the fifth resonance (219 Hz). It is known from the former distribution that an appreciable peak is formed near the bell end. This peak position, which is located at the distance of 378 cm from the mouthpiece input, exactly corresponds to the input side of the narrow cylindrical tube involved in the stopping mute (see Fig. 10). In other words, a "closed-closed" pipe pattern of the sixth order is derived from a "closed-open" pipe pattern of the fifth order when the stopping mute causes a semi-tone ascent.

The internal pressure distribution is also compared between the stopping mute and the hand-stopping model. The result is illustrated in Fig. 13 for the sixth resonance (cf. Tables 1 and 2). Note that the radiated pressure p_B at the bell output is supposed to be 1 Pa in the hand-stopping model and 5 Pa in the stopping mute for better visibility (actually, the radiated pressure in the stopping mute is larger than that in the hand-stopping model if the mouthpiece pressure is assumed to be the same). These two situations yield quite similar result except for the distribution around the final maximum near the bell. This acoustical difference seems to depend on the geometrical difference between the stopping mute (showing continuous change in inner geometry, particularly before and after the narrow cylindrical pipe) and the stopped-horn model (showing discontinuous change at the orifice of the modeled plate). Although a further improvement is suggested for the hand-stopping model, the result given in Fig. 13 confirms the validity of our simple modeling of the hand stopping.

3.3 Tonal Difference Between Stopping Mute and Hand Stopping

Although the stopping mute and hand stopping bring similar impression of the stopped tone, the corresponding acoustical characteristics are significantly different as shown in Fig. 1c, d. The muted-horn sound spectrum of Fig. 1d indicates the

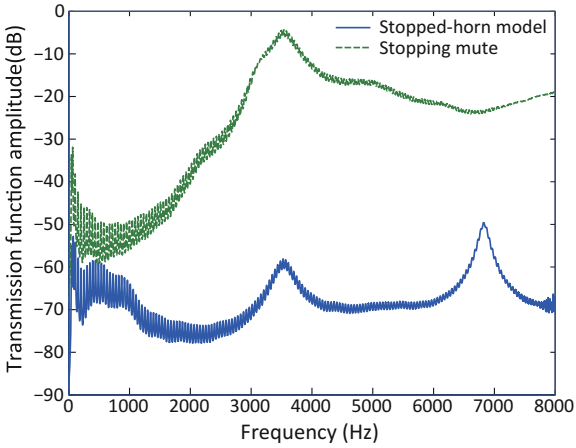


Fig. 14 Transmission function amplitude H (f) for the horn with the stopping mute and the stopped-horn model on the hand stopping, respectively

level depression in low frequencies below about 2 kHz and the level enhancement in higher frequencies above about 4 kHz. These changes might be attributed to the geometrical difference between the two illustrated in Fig. 10, which should bring about the difference in the transmission function amplitude H (f) defined as 20log $|p_B/p_M|$ from Eq. (5). This $H(f)$ is drawn in Fig. 14 for the stopping mute (the green dashed line) and the stopped-horn model on the hand stopping (the blue solid line [19]), respectively.

As shown in Fig. 10, the stopped-horn model has discontinuous change in cross section at both ends of a short tube (25 mm in length and 6 mm in diameter), while the stopping mute has smoother change in cross section at both ends of a short tube (41.5 mm in length and 8 mm in diameter). These geometrical differences might cause significant differences in the magnitude of H (f) between the both. Particularly, H (f) of the stopping mute indicates the characteristic of high-pass filter above 3.5 kHz, which seems to yield the spectrum-level enhancement shown in Fig. 1d. Also, H (f) of the stopping mute indicates much lower amplitudes below 2 kHz, which seem to yield the spectrum-level depression shown in Fig. 1d compared with Fig. 1c. These major differences in H (f) possibly produce tonal differences between the horn equipped with the stopping mute and the horn played with the hand stopping.

The resonance peaks near 3.5 kHz in Fig. 14 are probably due to the mouthpiece. The same resonance peaks appear in H (f) of the open, normal (hand-in-bell), and hand-stopped horns [19]. Also, the peaks at 3.5 kHz shifted to 3.4 kHz when the volume of the mouthpiece cup was enlarged by about 20 % with the rim diameter fixed [19]. Another relatively high peak at 6.8 kHz of the stopped-horn model is due to the above-mentioned short tube (or the orifice in the modeled plate). Since the actual hand-stopped horn probably does not bring any high-pass filter characteristic, the spectrum level of its radiated sound gradually decreases above 4 kHz as shown in Fig. 1c.

4 Straight Mute for the French Horn

A typical mute for the French horn is the straight mute. In this section its modeling will be discussed.

4.1 Structure of Straight Mute

Two kinds of straight mutes are shown in Fig. 15. Many straight mutes have a double inside tube for fine tuning by sliding the innermost tube to adjust the tube length. The smaller end of the conical mute is open, and the larger end is closed. A realistic but a little simplified inner geometry of the straight mute is illustrated in Fig. 16a, where the adjustable tube length is fixed to 10.0 cm. Also, the left end diameter is 2.75 cm, and the right end diameter is 12.75 cm. The whole length of the mute is 24.42 cm. The diameter of the inside tube is 3.81 cm. The inside tube changes both its own air mass given by the tube length and the cavity inside the

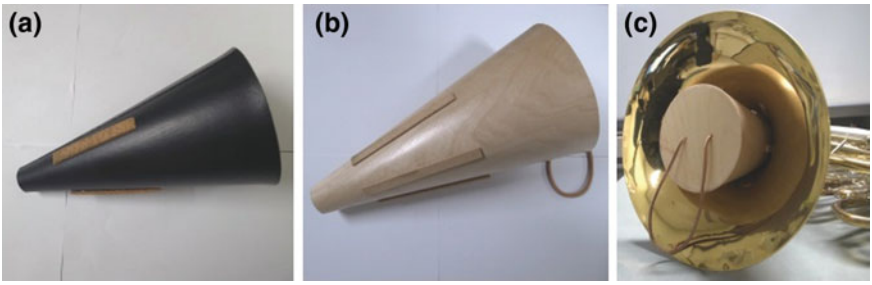


Fig. 15 Straight mutes (Yupon make) for the horn. **a** Made of fiber body and wooden bottom plate; **b** made of wood; **c** a wooden mute inserted into the horn bell

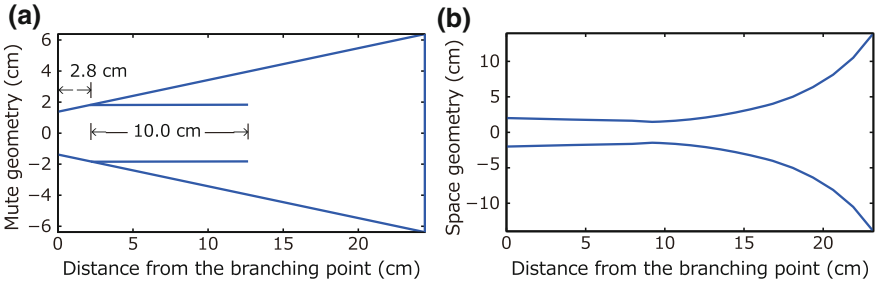


Fig. 16 Acoustical structures brought by the straight mute. **a** Realistic but a little simplified inner geometry of the mute with the inside tuning tube; **b** a tube equivalent to the annular spacing formed between the bell and mute [27]

conical mute. As a result, the mute itself forms the Helmholtz resonator whose frequency is slightly changed by the adjustable inside tube. Also, cork strips on the mute exterior firmly attach to the inner wall of the bell and then small opening is formed between the bell and mute. This opening with an annular shape may be transformed to an equivalent tube with the reduced radius distribution along the mute as shown in Fig. 16b in order to make numerical calculation possible.

A *branching system* consisting of the mute itself and the tube equivalent to the annular opening is thus made up when the mute is firmly inserted in the horn. The branching point is the smaller open end of the mute. Acoustical modeling of this branching system, which is the prime objective of our chapter, will be described in subsequent sections.

4.2 *Branching System Theory and Its Incorporation into* T*-Matrix Formulation*

A variety of acoustical systems have branches (dividing paths). An important example in woodwind instruments is the tone hole or finger hole. We have a long and extensive history on acoustical tone-hole research [9, 28–32], but its application to brass mutes seems to be inappropriate because of their much larger area and volume. Therefore, simpler branching-system theory [33, 34] may be a better choice for our brass-mute analysis, and its incorporation into the *T*-matrix representation of the muted-horn system will be treated.

Let us consider the simplest branching tube, where "Pipe 1" branches into "Pipe 2" and "Pipe 3" as depicted in Fig. 17. The quantities p_i and q_i (i = 1, 2, and 3) denote the acoustic pressures and acoustic volume velocities at the branching point, respectively. The following relations hold at the branching point:

$$\begin{cases} p_1 = p_2 = p_3 \\ q_1 = q_2 + q_3 \end{cases} \tag{9}$$

Therefore, the following fundamental admittance relation is given:

$$Y_1 = Y_2 + Y_3, \tag{10}$$

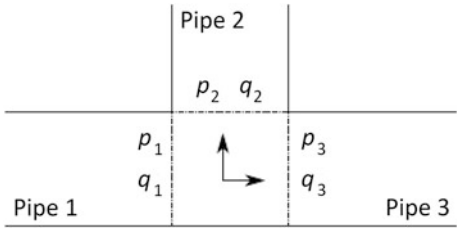


Fig. 17 The simplest branching system. "Pipe 1" branches into "Pipe 2" and "Pipe 3" [27]

where Y_1 is the output admittance of "Pipe 1", and Y_2 and Y_3 are the input admittances of "Pipe 2" and "Pipe 3", respectively. If a pipe branches into n pipes, Eq. (10) is generalized to $Y_1 = Y_2 + Y_3 + \cdots + Y_{n+1}$.

It is our next step to incorporate the admittance relation of Eq. (10) into the T-matrix formulation represented by Eq. (3). From Eq. (9) we have

$$\begin{pmatrix} p_1 \\ q_1 \end{pmatrix} = \begin{pmatrix} p_3 \\ q_2 + q_3 \end{pmatrix} = \begin{pmatrix} p_3 \\ p_2 Y_2 + q_3 \end{pmatrix} = \begin{pmatrix} p_3 \\ p_3 Y_2 + q_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ Y_2 & 1 \end{pmatrix} \begin{pmatrix} p_3 \\ q_3 \end{pmatrix} \tag{11}$$

Hence, it is possible to formulate our muted-horn system by incorporating the following branching matrix

$$T_{BR} = \begin{pmatrix} 1 & 0 \\ Y_2 & 1 \end{pmatrix} \tag{12}$$

into the total T-matrix at the mute input end. The matrix of Eq. (12) is defined by using the shunt admittance Y by Lampton [34].

In our case of the muted-horn system, it should be reasonable to consider that the mute itself is the branched system because the sound radiates from the annular spacing formed between the bell and mute. The Y_2 in the branching matrix of Eq. (12) can be calculated based on the geometry of the mute.

4.3 Acoustical Modeling of the Horn with the Straight Mute

If the inside tube does not exist in the mute, the horn with the mute is simply modeled as shown in Fig. 18a, where the mute inner cavity (Y_2) is considered as a branching system and the annular spacing (Y_3) is considered to be continuous from the horn body for sound radiation. The Y_2 is calculated by dividing the mute inner cavity into conical tubes (for convenience, the right figure is drawn by a cylindrical tube); the Y_3 by dividing the tube equivalent to the annular spacing shown in Fig. 16b into cylindrical and conical tubes. The calculated input impedance of this simplified model is illustrated in Fig. 19a in comparison with the open horn. It is clear that the mute raises resonance mode frequencies above the third mode.

If the inside tube does exists in the mute, acoustical situation becomes a little complicated. The input of the inside tube is the branching point between the tube itself and the annular spacing. Also, the output of the inside tube is the branching point between the mute cavity (③ in Fig. 18b) and another cavity formed between the mute and the inside tube (④ in Fig. 18b). The result of numerical calculation based on the branching system theory and the transmission matrix theory is

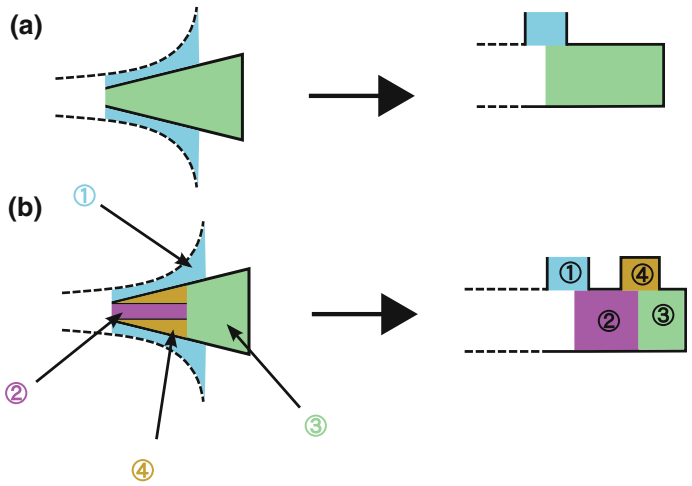


Fig. 18 Acoustical modeling of the straight mute. **a** A simplified model; **b** a realistic model with an inside tuning tube ②. These models are calculated by the branching tube theory which is incorporated into the transmission matrix formulation

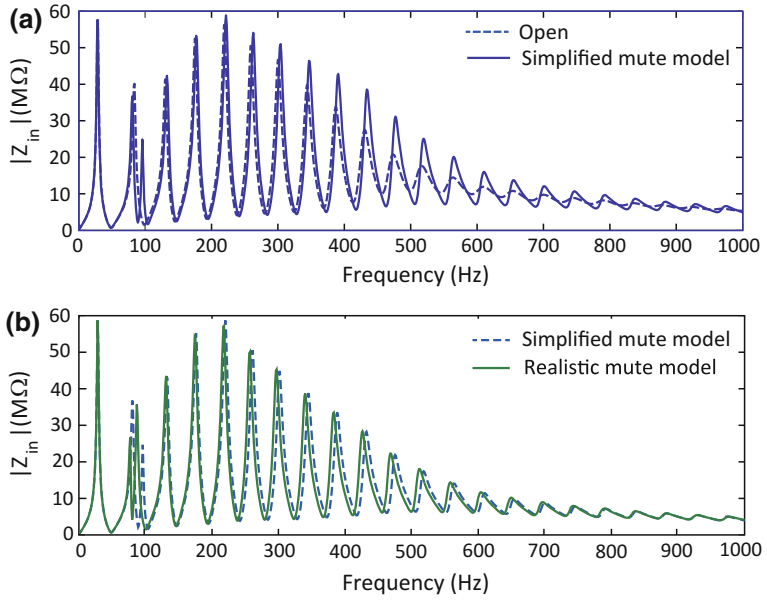


Fig. 19 The calculated input impedances of acoustical models of the straight mute for the horn (cf. Fig. 18). **a** A horn with a simplified straight-mute model versus an open horn; **b** a horn with a realistic straight-mute model versus a horn with a simplified straight-mute model

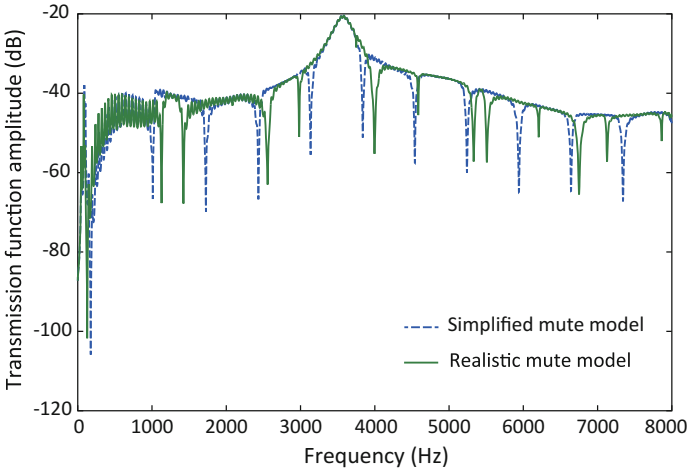


Fig. 20 Transmission functions of the straight-mute models for the French horn

illustrated in Fig. 19b. The simplified and realistic mute models denote the models without and with the inside tuning tube, respectively. It is well understood that the inside tuning tube reduces the mode frequencies. Therefore, if we consider Fig. 19a, b, we may infer that the inside tuning tube operates as a non-transposing device whose effect probably depends on the inertance of the inside tube and the diminished cavity of the mute interior. We can understand that the inside-tube length can adjust the pitch of the mode selected for the play. Such a pitch adjustment seems to be needed to the horn players because of their playing manner using their hand in the bell.

Next, the transmission function $H(f)$ defined by Eq. (5) is illustrated in Fig. 20, where the input pressure is the mouthpiece pressure and the output pressure is the bell pressure at the end of the annular spacing. A comparison between the simplified mute model (the blue dashed line) and the realistic mute model (the green solid line) is given in Fig. 20. A significant character brought by the mute is sharp dips, whose step is almost constant in the simplified model but is various in the realistic model. Particularly, the first dip around 80 Hz drastically falls down and this dip gives the Helmholtz resonance of the mute cavity. Higher dips above about 1 kHz probably correspond to higher Helmholtz or air-column resonances. Also, $H(f)$ between the first and second dips (corresponding to the horn resonances) rapidly changes with the amplitude difference of about 10 dB. As a whole the straight mute seems to form a high-pass filter. This result confirms the comments by Ancell [2] and Fletcher [9].

4.4 *Effects of Other Parameters of the Straight Mute*

Since Smith [4] suggested minor effects of other mute parameters on pitch improvement as well as on sound, we tried to calculate the effects of them. Smith [4] picked up the following parameters: (1) the thickness of the mute's corks, (2) a hole in the bottom plate of the mute, and (3) a detachable plug placed at the mute's top opening,.

1. **Cork thickness**
   If the mute's corks are thicker, the insertion depth of the mute should be shallower. The difference in the insertion depth between thin and thick corks is depicted in Fig. 21a. The cork thickness thus changes the branching point between the mute top and the annular spacing for sound radiation. As a result, the shape and length of the annular spacing are slightly changed. When the cork is thicker and the branching point is moved toward the bell end by 28 mm, the input impedance $|Z_{in}|$ (f) and the transmission function amplitude H (f) are affected as shown in Fig. 22. The thicker cork significantly changes the second mode and raises the higher mode frequencies (above the fifth mode). Also, the thicker cork appreciably raises H (f) in the playing frequency range below 1000 Hz.

2. **Bottom plate hole**
   If a hole is made in the bottom plate of the mute, the hole inertance can slightly reduce the resonance frequencies of the muted horn, particularly in lower modes. An acoustical model of the horn muted by the straight mute with a bottom hole is shown in Fig. 21b. The calculated Z_{in} (f) for the hole diameter of

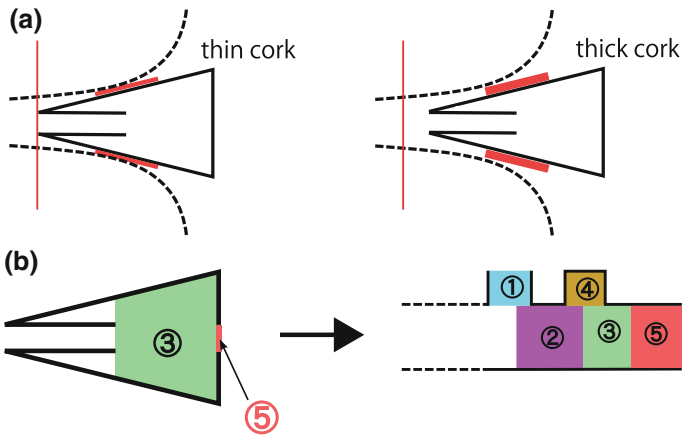


Fig. 21 Other structural parameters of the straight mute. **a** Thickness of corks; **b** a hole (denoted by ⑤) made in the bottom plate (cf. Fig. 18b)

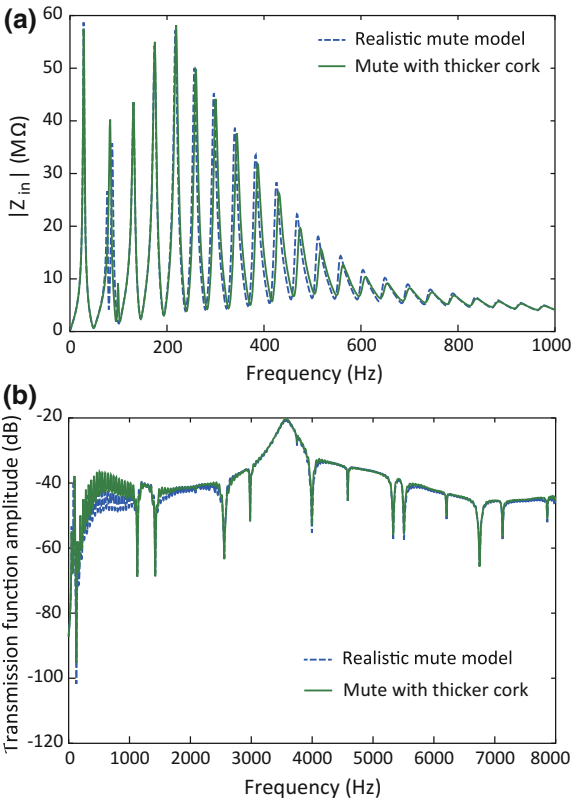


Fig. 22 The effects of the mute's corks on the input impedance $|Z_{in}|$ (f) and the transmission function amplitude H (f) when the cork is thicker and the branching point is moved toward the bell end by 28 mm

3/16 in. (=4.7625 mm) and the bottom plate thickness of 5 mm actually shows the above tendency. The peak frequencies of Z_{in} (f) with a bottom hole are reduced by about 10 cents around the 5th resonance mode. The transmission function amplitude H (f) is compared between the realistic model and that with a bottom hole (see Fig. 23). If a bottom hole is perforated, the dips of H (f) become shallow. This result might suggest the weakened capability of the energy absorption by the mute.

3. **Detachable top plug**

If a detachable plug is placed at the top opening of the straight mute, the top opening area is decreased and then the resonance frequencies of the muted horn are possibly diminished just as the case of the bottom plate hole. Our numerical calculation indicates that the detachable top plug causes the effects very similar to those by the bottom hole.

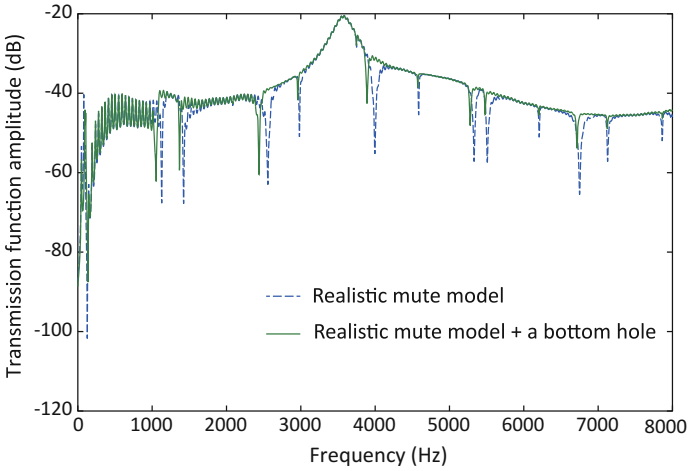


Fig. 23 The transmission function amplitude $H(f)$ for the realistic mute model and that for the realistic model added a hole made in the bottom plate of the mute

5 Application to Trumpet Mutes

There are various mutes for the trumpet. Bertsch [35] picked up 13 mutes (including the practice mute) and he investigated the influence of six common mutes on the dynamics, timbre, intonation, acoustic radiation, and frequency response of trumpets. In this section we will consider the three most-used mutes: the straight, cup, and wah-wah mutes. Numerical calculations based on their acoustical models are carried out by applying the branching system theory described in Sect. 4.2.

5.1 Structures and Models of Trumpet Mutes

Two kinds of straight mutes are shown in Fig. 24. One is aluminum made, another is cupper made. Trumpet straight mutes look similar to French horn ones shown in Fig. 15, although the part outside the bell is bowl-like. They have no inside tube for fine tuning in horn ones. Therefore, their basic structure is modeled by Fig. 18a for a simplified model of the horn straight mute.

A cup mute is shown in Fig. 25. This mute covers a bowl-like end of the straight mute with a cup, and the annular spacing formed between the mute cup and the bell tip is much reduced as shown in Fig. 25b. Then, the cup mute is modeled as depicted in Fig. 27a, where the cup adds a small cavity (brown part) and an opening (orange part) to the annular spacing (blue part) of the straight mute [27]. In other words, the output of the annular spacing branches to the inside cavity (tube) of the cup and the reduced opening (tube) for sound radiation. That is, the input and

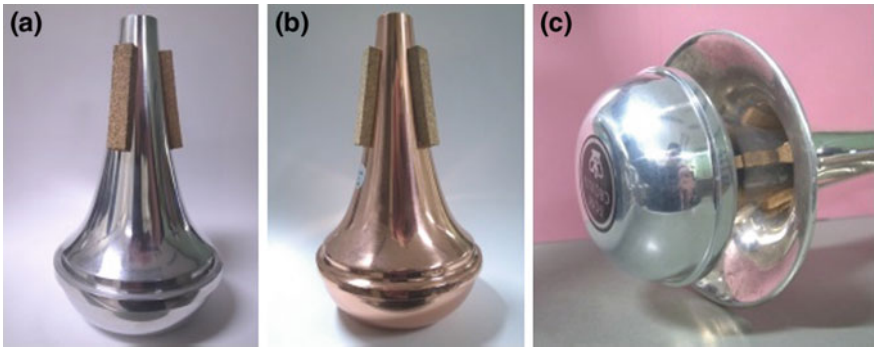


Fig. 24 Straight mute for the trumpet. **a** Aluminum made; **b** copper made; **c** external view when the mute is inserted in the trumpet

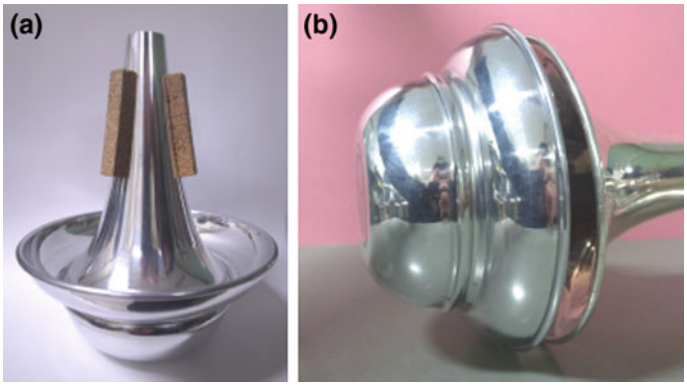


Fig. 25 Cup mute for the trumpet. **a** Total view; **b** external view when the mute is inserted in the trumpet

output of the annular spacing are the first and second branching points of the cup mute, respectively.

A trumpet wah-wah mute is shown in Fig. 26. The mute outer (left) end is open as shown in Fig. 26a, c. As a result, the so-called wah-wah tone is produced by moving the player's hand in front of this opening during the play. Also, the inside tube can be removed as indicated in Fig. 26b. Such a mute is called the Harmon mute. Since the wah-wah mute is tightly fitted to the bell wall at the cork position, there is no annular space between the mute and bell. The sound is radiated from the outer opening of the inside tube. Hence, the wah-wah mute is modeled as depicted in Fig. 27b, which is basically the same as the straight mute modeled as shown in Fig. 18a except for the cork part (pink part) that seems to be important when the inside tube is removed [27].

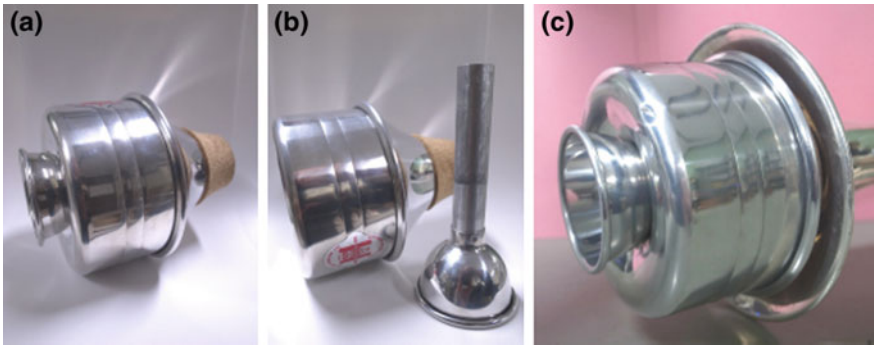


Fig. 26 Wah-wah (or wow-wow) mute for the trumpet. **a** Total view; **b** an inside tube removed; **c** external view when the mute is inserted in the trumpet

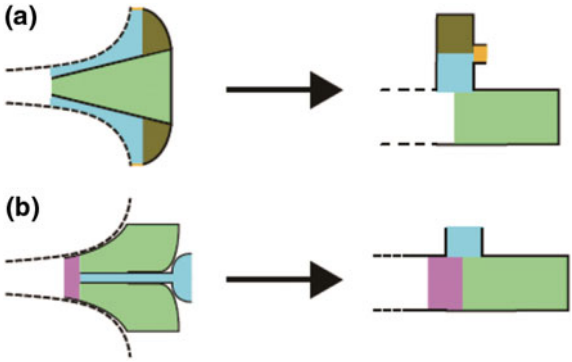


Fig. 27 Basic structures (*left*) and models (*right*) of trumpet mutes [27]. **a** The cup mute; **b** the wah-wah mute

5.2 *Numerical Calculation of the Trumpet with the Straight Mute*

The bore geometry of the B^b trumpet (Yamaha YTR-2320E) used for our numerical calculation is drawn in Fig. 28 [36]. Its total length is 137.1 cm (a flaring bell part: 56.2 cm; a cylindrical part: 49.0 cm; a conical lead pipe: 23.1 cm; a mouthpiece part: 8.8 cm). Numerical calculations on this trumpet muted by straight, cup, and wah-wah mutes are carried out below.

The straight mute shown in Fig. 24 has its inside geometry depicted in Fig. 29a and it brings the annular spacing between the bell wall and the mute, which is reduced to a bore geometry given in Fig. 29b. The outer ends of the mute and the annular spacing are closed and open, respectively. The diameter of the spacing end

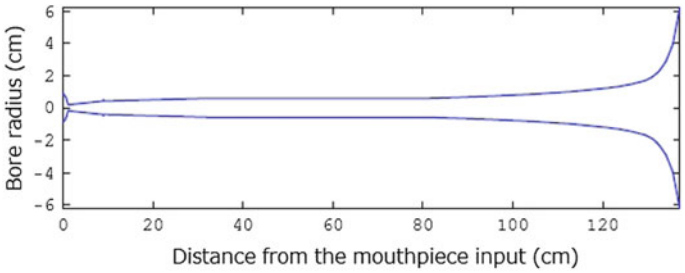


Fig. 28 Bore geometry of the B^b trumpet used for the calculation [36]

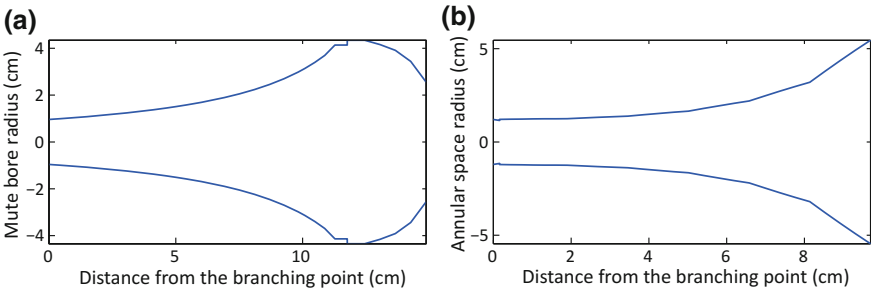


Fig. 29 The inside geometry of the trumpet straight mute (**a**) and the reduced bore geometry of the annular spacing between the bell wall and the mute (**b**)

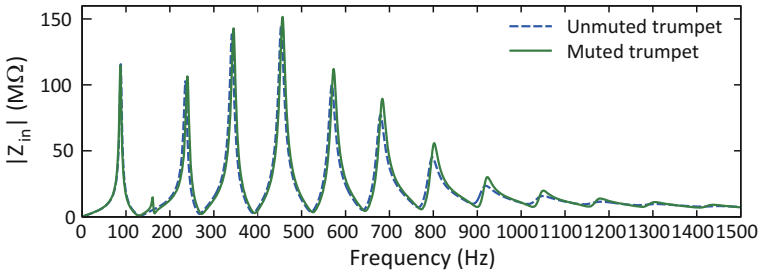


Fig. 30 Input impedances of the trumpet with and without the straight mute [27]

is 10.9 cm and the area for sound radiation is decreased to 77.8 %. The branching point is set to be at 127.4 cm from the mouthpiece input.

The calculated input impedance of the trumpet equipped with the straight mute is shown in Fig. 30 in comparison with that of the unmuted trumpet. Although the effects of the straight mute are not so significant compared with the horn case, we can recognize (1) the appearance of a new peak between the original first and second peaks, (2) an overall shift of peak frequencies to slightly higher frequencies, and (3) an appreciable increase of peak amplitudes in high frequency range. The

peak frequencies shift by around 4 Hz, which corresponds to about 35 cents at a peak of 240 Hz and about 6 cents at a peak of 923 Hz. Differences in the mute effects between the horn and trumpet cases should be investigated in the future.

The transmission function of the muted trumpet is shown in Fig. 31 in comparison with that of the unmuted trumpet. We can recognize sharp dips in steps of almost constant frequency (around 1050 Hz) for the muted trumpet. These dips are observed in the muted horn (not shown in this chapter) [27]. They are derived from the Helmholtz resonances (dips at 223 and 1758 Hz) and the air-column resonances (dips at 2805, 3777, 4822, 5895 Hz, and so on) of the mute interior. Small peaks with regular spacing correspond to the muted-trumpet resonances.

The input impedance and transmission function of the annular spacing respectively shown in Fig. 32a, c suggest that the straight mute operates as a high-pass filter. This result may be coincident with the description of Backus [3]: The straight mute for the trumpet operates as a high-pass filter for the frequencies above 1800 Hz. The input impedance being composite at the branching point is shown in Fig. 32b. A sharp peak is clearly recognized at 165 Hz. Since this peak is below B^{b}_{3} (233 Hz), it does not bring any undesirable effects to the play [3]. Probably, this peak at 165 Hz has a significant relation with a new peak (at 161 Hz) found in the input impedance of the muted trumpet shown in Fig. 29.

Finally, examples of acoustic pressure distributions along the bore are shown in Fig. 33a, b, respectively. The former is on the fourth mode (the fourth peak of $|Z_{\mathrm{in}}|$, 453 Hz) of the unmuted trumpet; the latter on the fourth mode (the fifth peak of $|Z_{\mathrm{in}}|$, 457 Hz) of the trumpet with the straight mute (cf. Fig. 29). Although p_{B} at the bell end is assumed to be 1 Pa for the both, the muted trumpet indicates the internal

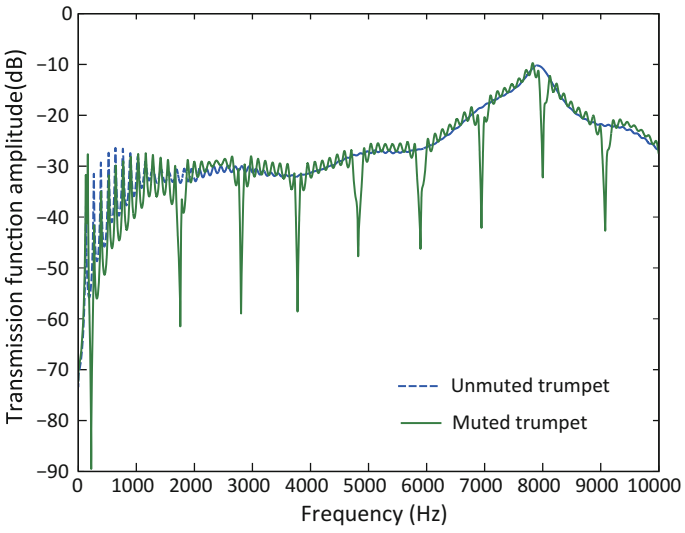


Fig. 31 Transmission function amplitude $H(f)$ of the trumpet with and without the straight mute [27]

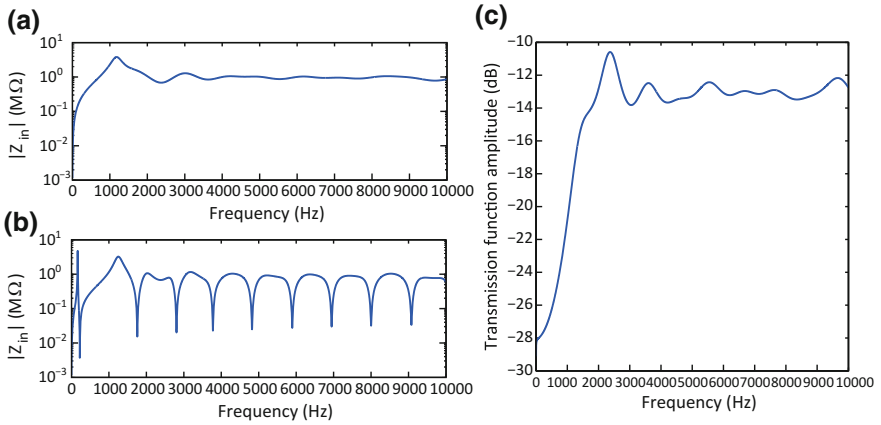


Fig. 32 The input impedance (**a**) and transmission function amplitude (**c**) of the annular spacing, and the input impedance being composite at the branching point (**b**) [27]

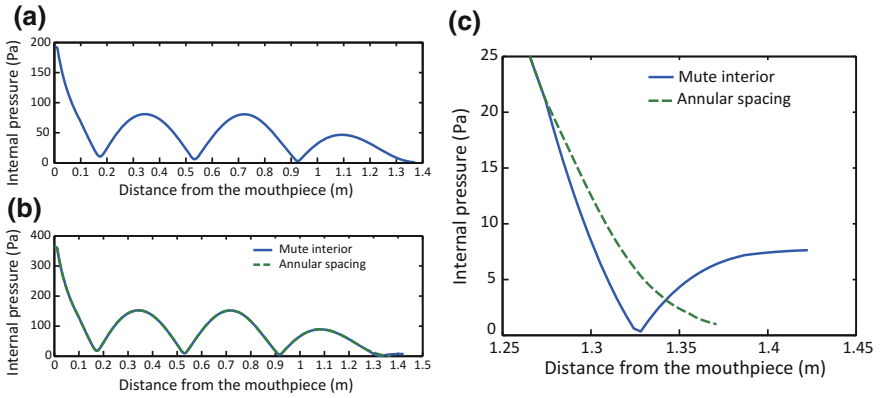


Fig. 33 Acoustic pressure distributions along the bore (p_B = 1 Pa) [27]. **a** The fourth mode of the unmuted trumpet (453 Hz); **b** the fourth mode (the fifth peak of the input impedance) of the trumpet with the straight mute (457 Hz); **c** the enlargement around the end portion of the internal pressures along the trumpet with the straight mute

pressure about twice as large as that of the unmuted trumpet. In other words, the muted trumpet should be blown harder to produce the sound radiation with the same amplitude. Also, the internal pressures near the end portion of the muted trumpet are compared between the mute interior (the solid line) and the annular spacing between the bell wall and the mute (the dashed line) in Fig. 33c. The acoustic pressure at the closed end of the straight mute is about eight times the radiated pressure at the bell end.

5.3 *Numerical Calculation of the Trumpet with the Cup Mute*

When the trumpet is equipped with the cup mute, the reduced bore shapes of the modeled elements (see Fig. 27a) are depicted in Fig. 34. The right ends of the mute interior (a) and the cup interior (c) are closed. The tube diameter corresponding to the radiating area (d) is 8.3 cm and the tube length is given by the open end correction. This area is 45 % of the unmuted trumpet. The first branching point A at the input of the annular spacing (b) is set to be 127.4 cm from the mouthpiece and the second branching point B at the output of the annular spacing is set at the cup tip [27].

The input impedance of the trumpet with the cup mute is shown in Fig. 35, which is very similar to Fig. 30 on the trumpet with the straight mute. This is probably because the mute interior and the annular spacing between the bell wall and mute are almost the same between the both (cf. Figs. 29 and 34). Backus [3] says that the effects of the cup mute are quite similar to those of the straight mute within the playing frequency range.

On the other hand, the transmission function amplitude H (f) of the cup mute shown in Fig. 36 looks quite different from that of the straight mute shown in Fig. 31. This difference is due to the cup interior (shown in Fig. 34c), which brings the dips at 2315, 6550, and 8810 Hz in Fig. 36. These three dips completely

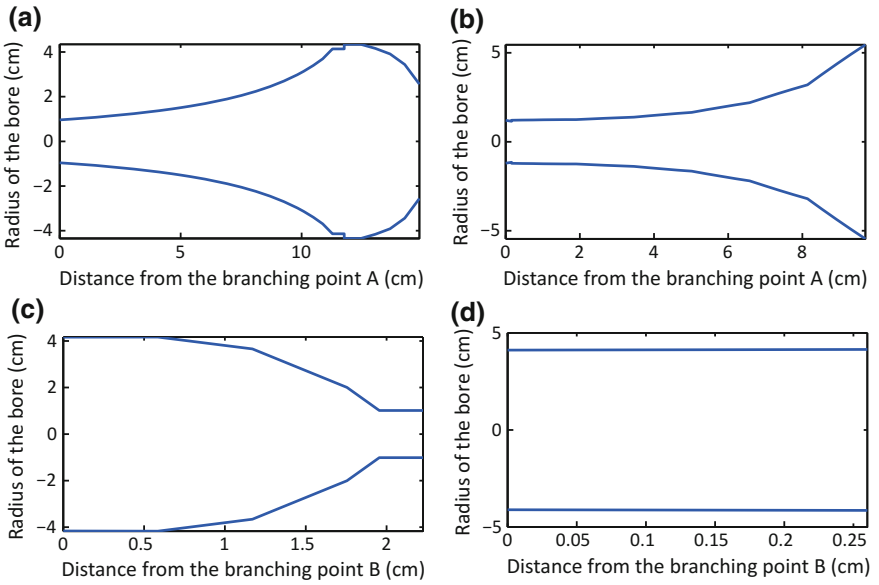


Fig. 34 The reduce bore geometries of the modeled elements when the cup mute is inserted in the trumpet [27]. **a** Mute interior; **b** annular spacing between the bell wall and mute; **c** cup interior; **d** radiating area. Elements (**a**) and (**b**) are branched at the first branching point A; elements (**c**) and (**d**) are branched at the second branching point B

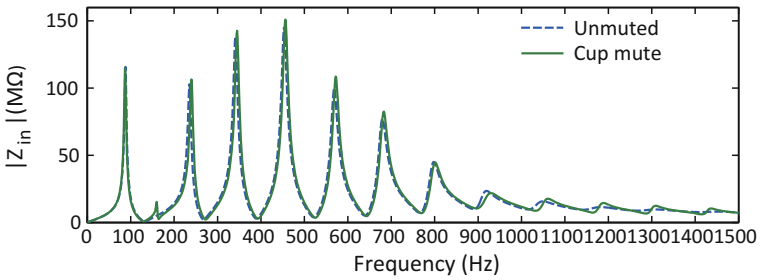


Fig. 35 Input impedances of the trumpet with and without the cup mute [27]

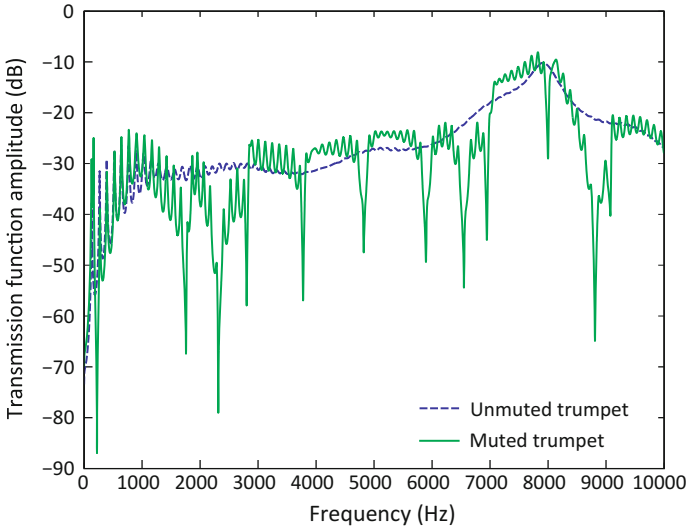


Fig. 36 Transmission function amplitude H (f) of the trumpet with and without the cup mute [27]

correspond to the impedance minima of the cup interior itself [27]. Also, they appear in Fig. 37a, which indicates the composite impedance of the cup and the open area branching at the radiation end. If these three dips are removed from Fig. 36, the remaining nine dips almost exactly matches those in Fig. 31.

The cup mute yields higher level of H (f) between these nine dips. This is probably due to the branched paths from the mute input to the cup interior and to the opening for the radiation. The transmission function amplitude seen from the latter branched path shown in Fig. 37c indicates the three dips of cup resonances and the wavy changes between the dips due to the annular-space resonances. These weak resonances appear in Fig. 37b, too. Also, it should be paid attention toward three pairs consisting of a dip and a peak at around 2.3, 6.7, and 9.0 kHz (marked by the red dashed line). These may be considered as a kind of mode repelling that is

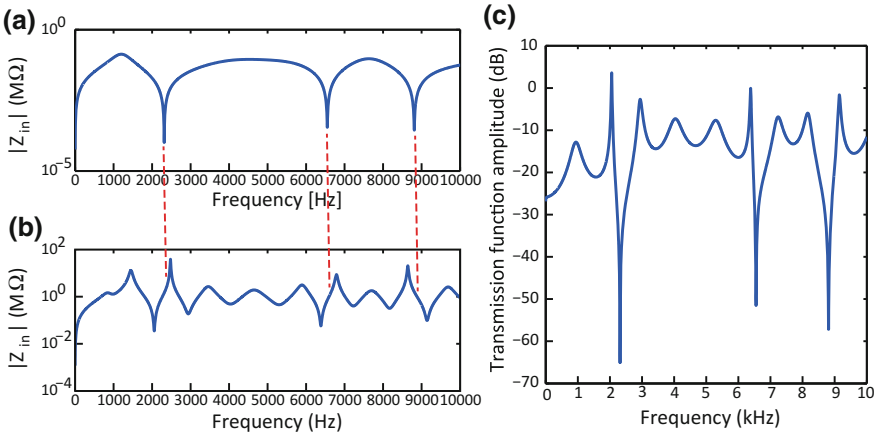


Fig. 37 Input impedances and transmission function amplitude of the construction elements of the cup mute (cf. Fig. 34). **a** The composite impedance of the cup cavity and the radiating open area seen at the second branching point B (at the output of the annular spacing); **b** the input impedance seen from the input of the annular spacing; **c** the transmission function amplitude seen from the input of the annular spacing

caused if and when the resonance frequencies of two acoustical systems are sufficiently close to each other [9]. In Fig. 37b both resonances of the annular spacing and the cup cavity are involved.

Particularly, the difference from the straight mute is appreciable between about 700 Hz and about 1400 Hz in Fig. 36. The cup mute may be then recognized as a band-pass filter for this frequency band, whose center frequency (about 900–1100 Hz) is produced by the resonance formed between the cup cavity and the opening inertance [3].

The resonance characteristics of the mute interior cavity are shown in Fig. 38a. Moreover, those of the mute cavity plus the annular spacing side (including the cup cavity and the radiating open area) are shown in Fig. 38b. We may recognize the maxima and minima of $|Z_{in}|$ in Fig. 38a as those of $|Z_{in}|$ in Fig. 38b as marked by the red dashed line. The changes between those are due to the annular spacing (cf. Fig. 37 c).

Finally, an example of internal pressure distribution along the bore is considered. Although the fourth mode (the fifth peak of the input impedance) of the trumpet with the cup mute indicated the resonance frequency of 456 Hz and the standing-wave pattern very similar to Fig. 33b, the mouthpiece pressure was only about 240 Pa compared to about 370 Pa in the case of the straight mute [27].

Since the radiating pressure at the end of the tube shown in Fig. 34d is 1 Pa, this reduction of the mouthpiece pressure should be the effect of the cup. The enlargement around the end portion is shown in Fig. 39. The pressure amplitudes are relatively low compared with the straight mute case shown in Fig. 33c. The acoustic pressure in the cup interior (shown by the red dashed line from the second

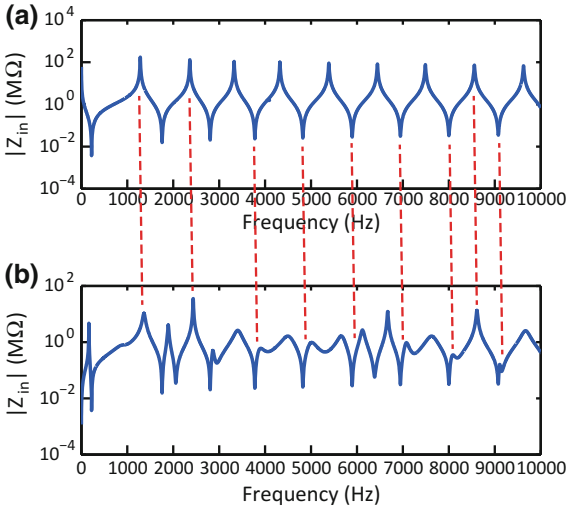


Fig. 38 Input impedances of the construction elements of the cup mute (cf. Fig. 34). **a** The input impedance of the mute interior; **b** the composite input impedance of the mute interior and the annular spacing side seen from the first branching point A

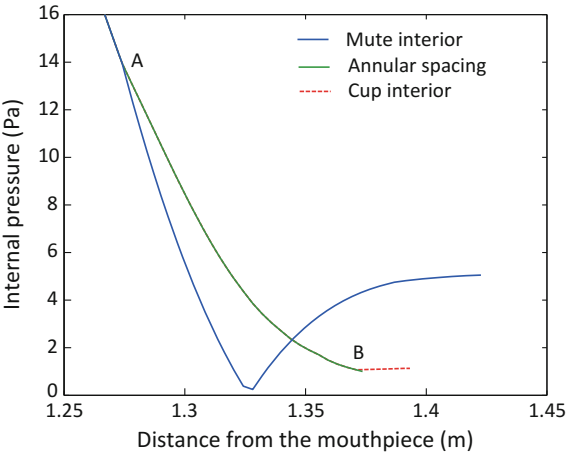


Fig. 39 The enlargement around the end portion of the internal pressures along the trumpet with the cup mute. The mute interior and the annular spacing are branched at point A; the annular spacing and the cup interior are branched at point B

branching point B) slightly increases toward the cup end. The acoustic pressure at the end of the mute interior (shown by the blue solid line) is about three times that at the end of the cup interior.

5.4 Numerical Calculation of the Trumpet with the Wah-Wah Mute

As shown in Fig. 27b, the wah-wah mute branches into the mute interior cavity and the inside tube for sound radiation. The mute interior cavity and the inside tube are

reduced to the bores shown in Fig. 40a, b, respectively. The end of the mute interior is closed, and the diameter at the end of the inside tube is 5.45 cm. The radiating area is 14.5 % of the unmuted trumpet. The inside tube length is 11.3 cm. The branching point is set to be 131.8 cm from the mouthpiece input.

The input impedances of the trumpet with and without the wah-wah mute are shown in Fig. 41. The logarithmic representation is given in Fig. 41b for the comparison with Backus [3]. There appear two clear peaks below and above the first mode of the unmuted trumpet. The latter peak (at 103 Hz) might be a "new

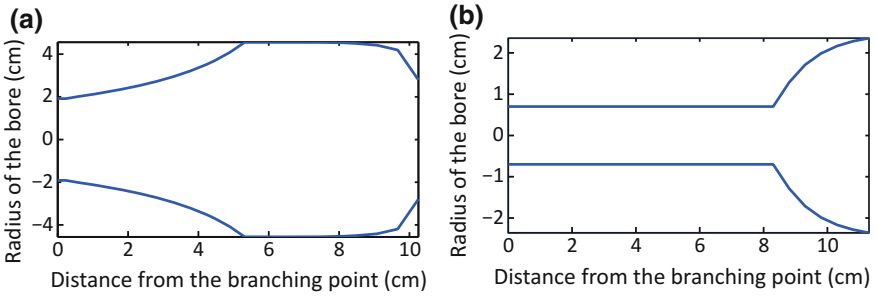


Fig. 40 The reduced bore shapes of the mute interior cavity (**a**) and the inside tube (**b**) of the wah-wah mute for the trumpet [27]

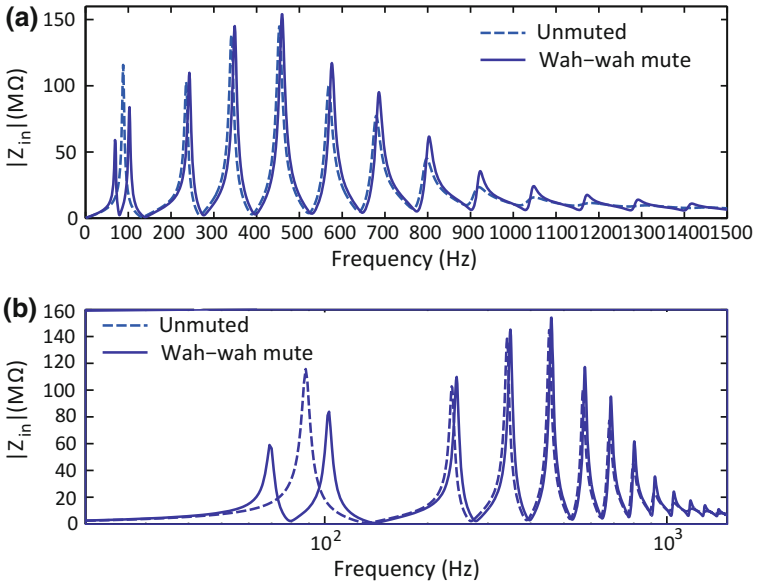


Fig. 41 The input impedances of the trumpet with and without the wah-wah mute. **a** Linear scale of the frequency; **b** logarithmic scale of the frequency [27]

peak" appearing between the first and second peaks of the unmuted trumpet, but its level is quite high and the frequency is very close to the first peak of the unmuted trumpet. Although its reason is not specified, the resonance of the mute cavity might be involved. The experimental result of Backus [3] is almost coincident with our result shown in Fig. 41b except for the first peak higher than the second peak. Other characteristics shown in Fig. 41 are very similar to those brought by the straight and cup mutes.

The transmission function amplitude of the wah-wah mute is given in Fig. 42. We may recognize the dips (at 450, 2159, 3580, 5011, 6478, and 9425 Hz) appearing in steps of almost constant frequency (about 1.5 kHz) as in the case of the straight mute, which gave about 1.05 kHz for the constant frequency step (see Fig. 31). The dip frequencies above correspond to the frequencies at the impedance minima of the mute interior as in the straight mute (cf. Fig. 27). However, the wah-wah mute indicates the characteristics different from the straight mute below and above around 1.5 kHz. The H (f) of the wah-wah mute is largely diminished below 1.5 kHz, however, it shows the increased levels between the dips. The wah-wah mute may be recognized as a band-pass filter between 1.5 and 2 kHz [3].

Finally the calculation result on the internal pressure distribution is shown in Fig. 43, where the end portion is enlarged for the fourth mode (at 460 Hz) of the trumpet with the wah-wah mute (the radiated pressure at the output of the inside tube is assumed to be 1 Pa). Although the mouthpiece pressure of the unmuted trumpet was about 200 Pa as shown in Fig. 33a, that of the muted trumpet with the wah-wah mute was about 3×10^4 Pa. Also, the pressure at the mute interior end is around 800 times the radiated pressure. These effects seem to be due to the mode

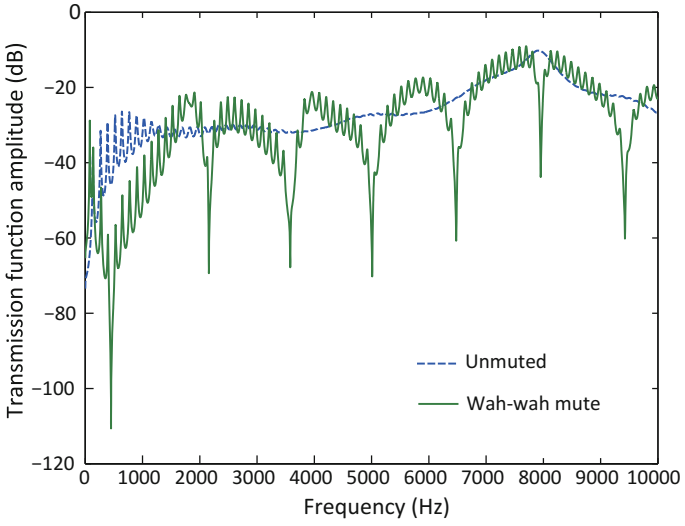


Fig. 42 Transmission function amplitude H (f) of the trumpet with and without the wah-wah mute [27]

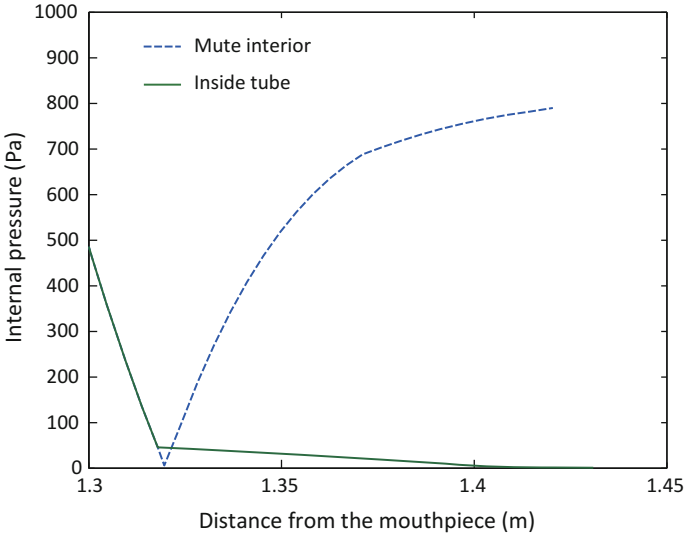


Fig. 43 The enlargement around the end portion of the internal pressures along the trumpet with the wah-wah mute. The fourth mode (at 460 Hz) is drawn by assuming the radiated pressure at the output of the inside tube to be 1 Pa. The mute interior and the inside tube are branched at 131.8 cm distant from the mouthpiece input

frequency of 460 Hz close to 450 Hz of the dip frequency of $H(f)$ in Fig. 42, which corresponds to the Helmholtz resonance frequency of the mute interior cavity. Interestingly, since the pressure node is formed near the branching point, the internal pressure along the inside tube is much smaller than that in the mute interior. Thus, the wah-wah mute brings peculiar characteristics compared with the straight and cup mutes.

5.5 *Appearance of a New Peak in Muted-Brass Input Impedance*

It was recognized that new peaks always appear in the input impedances of the muted brasses: See Fig. 19a for the horn with the straight mute, Fig. 30 for the trumpet with the straight mute, Fig. 35 for the trumpet with the cup mute, and Fig. 41 for the trumpet with the wah-wah mute. When these three mutes are inserted into the brass, the brass bore commonly branches to the mute interior and to the annular spacing or the inside tube. Therefore, the composite impedance of these branching systems seems to affect the input impedance of the muted brass. The muted-brass input impedance calculated at the mouthpiece end and the composite input impedance calculated at the branching point are compared with each other in Fig. 44 for the four cases above. Also, a comparison of the frequency and

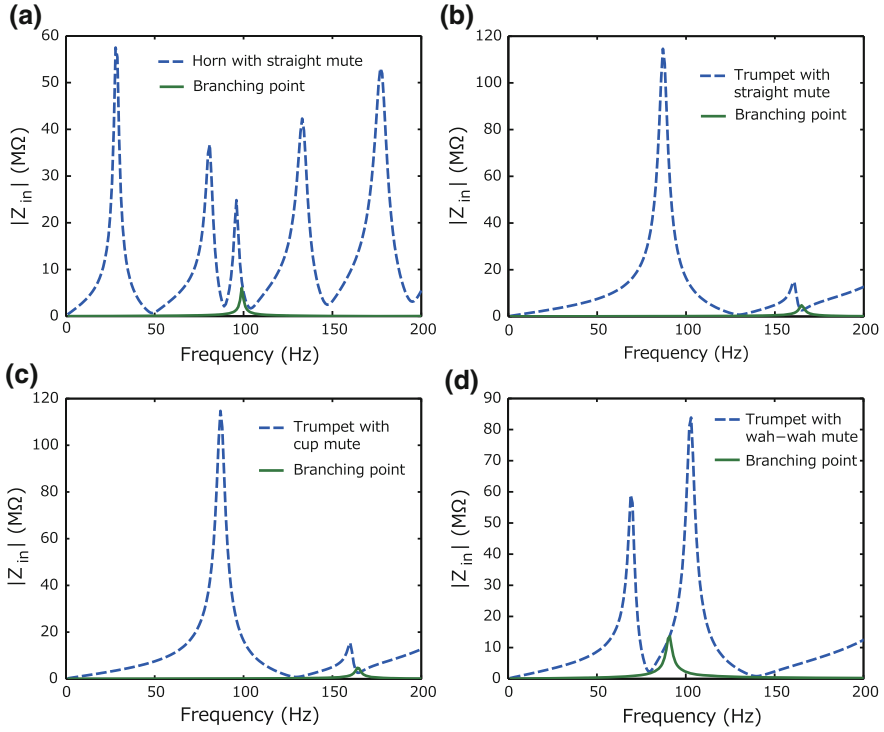


Fig. 44 The input impedance of the muted brass and the composite input impedance of the branching systems. **a** The horn with the straight mute; **b** the trumpet with the straight mute; **c** the trumpet with the cup mute; **d** the trumpet with the wah-wah mute. It should be noted that the simplified model without the inside tuning tube is used for the horn with the straight mute (cf. Figs. 18 and 19) for the comparison with the trumpet with the straight mute

magnitude of $|Z_{in}|$ (f) between a new peak added by the muted brass and a peak brought by the branching systems is given in Table 3.

Although the first-mode resonance magnitude given by the composite input impedance of the branching systems is quite small compared with the mode resonance magnitudes of the muted-brass input impedance, the frequency of the first-mode resonance given by the composite impedance is very close to the frequency of the new peak in the input impedance added by the mute except for the trumpet/wah-wah mute case. The appearance of the new peak in the muted-brass input impedance may be caused by the composite impedance of the branching systems. The difference in the frequency and magnitude between the new peak in the muted- brass input impedance and the first peak in the composite impedance possibly depends on the resonance characteristics of the brass bore.

As Backus [3] pointed out, the new peak in the muted-brass input impedance appears in low frequency range which is not used in usual play (in the muted trumpet, it appears between the first and second peaks). This implies that acoustical

Table 3 A comparison of the frequency and magnitude of $|Z_{in}|$ between a new peak (impedance maximum) added by the muted brass and a peak brought by the branching systems (the mute interior and the annular spacing or inside tube)

Brass/mute	Horn/straight	Trumpet/straight	Trumpet/cup	Trumpet/Wah-Wah
New peak frequency (Hz) in $\|Z_{in}\|$ of the muted brass	96	161	160	103
Peak frequency (Hz) in $\|Z_{in}\|$ of the branching systems	99	165	164	91
$\|Z_{in}\|$ value (MΩ) of the muted brass	24.9	14.7	15.4	83.9
$\|Z_{in}\|$ value (MΩ) of the branching systems	6.63	4.79	4.70	13.6

design of the mutes should be done so that this new peak never appears in the usual playing frequency range.

On the other hand, although the stopping mute for the horn yields a small new peak in the input impedance as shown in Fig. 11a, the nature of this new peak is different from that of the muted brass mentioned above. This is because the stopping mute has no branching point as indicated in Fig. 10. Hence, the seeming new peak in Fig. 11a might be only a small hump on the lower frequency side of the third resonance mode as Backus [3] suggested (cf. Sect. 3.1).

6 Conclusions

Several common mutes for brasses (the stopping and straight mutes for the horn; the straight, cup, and wah-wah mutes for the trumpet) have been considered in this chapter. Prior to the acoustical modeling of the muted brass, the effects of the player's right hand in the horn playing were investigated. Particularly, the mechanism of the hand stopping in the horn was extensively analyzed on the basis of the conventional transmission matrix (T-matrix) theory.

The calculated internal pressure distribution along the horn revealed that the hand for the hand stopping brought a large inertance and made the wavelength in the horn longer than the original one. As a result, the original mode (order n) in the "closed-open" pipe is shifted down to the next lower mode (order $n - 1$) in the "closed-closed" pipe. This shift is the essential cause of the puzzling pitch *descent* by hand stopping. However, if the comparison is made at the same mode order n, hand stopping produces pitch *ascent* of about a semitone. Moreover, the physical cause of *metallic* timbre by hand stopping was pursued by experimental methods: the immobilization experiment (by immersing the horn in the sand) indicated that wall vibration of the born bell and pipe was not the cause; the mouthpiece pressure measurement revealed that the nonlinear propagation along the bore should be the

major cause [19]. The metallic timbre is generated from the *wave corrugation* brought by the nonlinear propagation starting from the mouthpiece pressure. However, it is our future topic whether the same wave corrugation really occurs in the horn muted by the stopping mute or not.

The transmission function amplitude (the absolute ratio of the radiated pressure to the mouthpiece pressure) $H(f)$ for the horn with the stopping mute and the stopped-horn model on the hand stopping well explained significant differences in tones produced by the horn muted with the stopping mute and by the horn played with hand stopping. The hand stopping yielded a rather flat $H(f)$ on the whole, while the stopping mute yielded a high-pass filter characteristic above 3.5 kHz and a strong amplitude depression below 2 kHz.

Although the *T*-matrix theory is straightforwardly applied to the horn with the stopping mute and the hand-stopped-horn in order to calculate the acoustic quantities such as the input impedance, transmission function, and internal pressure distribution along the bore, it is difficult in considering other mutes. This is because when the mute is inserted in the brass, the brass bore branches to the mute interior and the annular spacing between the bell inner wall and the mute outer surface as in the case of the straight mute. Therefore, the branching system theory was incorporated into the *T*-matrix formulation for our acoustical modeling of the muted horn and trumpet.

Calculation results of the input impedance showed (1) the appearance of a new peak between the original first and second peaks, (2) an overall shift of peak frequencies to slightly higher frequencies, and (3) an appreciable increase of peak amplitudes in high frequency range when the straight, cup, and wah-wah mutes were used. The appearance of a new peak and an overall shift of peak frequencies may be caused by the composite input impedance of the branching systems. Also, since the branched annular spacing diminishes the radiation area and may increase the cutoff frequency, peak amplitudes tend to be raised in high frequency range.

Calculation results of the transmission function amplitude indicated sharp dips, which corresponded to the resonance frequencies of the mute interior cavity. Also, the transmission function of the annular spacing contributed to the amplitude between the dips. Calculation results of the internal pressure distribution indicated a significant increase of the mouthpiece pressure. This means that the enhancement of the playability as well as the suppression of the radiated sound pressure is brought by the mutes.

Our main focus was on the acoustical modeling of the straight mute for the French horn and the cup mute for the trumpet. These two cases are very interesting because the former usually has an inside tuning tube and forms a complicated branching systems in the mute interior, and the latter makes two branching points along the annular spacing branch for sound radiation. Since the inside tuning tube in the former reduced the mode (peak) frequencies of the input impedance, it could operate as a non-transposing device whose effect might depend on the inertance of the inside tube and the diminished cavity of the mute interior. Also, effects of other parameters (the thickness of the mute's cork, a hole perforated in the bottom plate of the mute, and a detachable plug placed at the mute's top opening) of the straight

mute were briefly considered. The latter indicated three sharp dips on the curve of the composite input impedance of the cup cavity and the radiating open area seen at the second branching point (at the output of the annular spacing). As a result, corresponding extra three dips were added to the transmission function amplitude of the trumpet equipped with the cup mute.

Bibliography

1. Baines, A.: The Oxford Companion to Musical Instruments, pp. 216–217. Oxford University Press, New York (1992)
2. Ancell, J.E.: Sound pressure spectra of a muted cornet. J. Acoust. Soc. Am. **32**, 1101–1104 (1960)
3. Backus, J.: Input impedance curves for the brass instruments. J. Acoust. Soc. Am. **60**, 470–480 (1976)
4. Smith, N.E.: The horn mute: an acoustical and historical study. Doctoral dissertation. Eastman School of Music, University of Rochester, USA (1980)
5. Watts, A.: Spectral analysis of the French horn and the hand-in-bell effect. Ph.D. dissertation. Department of Physics, University of Illinois at Urbana-Champaign, USA (2009)
6. Dell, N.L.: Investigation into factors affecting the acoustics and playability of the horn, and the effect of auditory feedback on the playing. Doctoral dissertation. Department of Music, University of Western Australia, Australia (2012)
7. Farkas, P.: The Art for French Horn Playing, pp. 11–14, 78–81. Indiana University Press, Indianapolis, USA (1962)
8. Tuckwell, B.: Playing the Horn, p. 15. Oxford University Press, New York, USA (1978)
9. Fletcher, N.H., Rossing, T.D.: The Physics of Musical Instruments, Chaps. 7, 8, and 14, 2nd edn. Spriner, New York, USA (1998)
10. Backus, J.: The Acoustical Foundations of Music, pp. 228–235. W. W. Norton, New York, USA (1969)
11. Meyer, J.: Acoustics and the Performance of Music, translated by U. Hansen, 5th edn., pp. 45–53. Springer, New York, USA (2009)
12. Morley-Pegge, R.: The French Horn, p. 137. Ernest Benn, London, UK (1960)
13. Hirschberg, A., Gilbert, J., Msallam, R., Wijnands, A.P.: Shock waves in trombone. J. Acoust. Soc. Am. **99**, 1754–1758 (1996)
14. Thompson, M.W., Strong, W.J.: Inclusion of wave steepening in a frequency-domain model of trombone sound production. J. Acoust. Soc. Am. **110**, 556–562 (2001)
15. Meyer, A., Pyle Jr., R.W., Gilbert, J., Campbell, D.M., Chick, J.P., Logie, S.: Effects of nonlinear sound propagation on the characteristic timbres of brass instruments. J. Acoust. Soc. Am. **131**, 678–688 (2012)
16. Whitehouse, J.: A study of the wall vibrations excited during the playing of lip-reed instruments. Ph.D. thesis. Technology Faculty, The Open University, (2003)
17. Moore, T.R., Shirley, E.T., Codrey, I.E.W., Daniels, A.E.: The effects of bell vibrations on the sound of the modern trumpet. Acta Acust. Acust. **91**, 578–589 (2005)
18. Kausel, W., Zietow, D.W., Moore, T.R.: Influence of wall vibration on the sound of brass wind instruments. J. Acoust. Soc. Am. **128**, 3161–3174 (2010)
19. Ebihara, T., Yoshikawa, S.: Nonlinear effects contributing to hand-stopping tones in a horn. J. Acoust. Soc. Am. **133**, 3094–3106 (2013)
20. Dell, N., James, R., Davidson, J., Wolfe, J.L.: The effect of hand and mute on the impedance spectra of the horn. In: Proceedings of International Symposium on Musical Acoustics, pp. 1–5, Sydney, Australia (2010)

21. Caussé, R., Kergomard, J., Lurton, X.: Input impedance of brass musical instruments: comparison between experiment and numerical models. J. Acoust. Soc. Am. **75**, 241–254 (1984)
22. Mapes-Riordan, D.: Horn modeling with conical and cylindrical transmission-line elements. J. Audio Eng. Soc. **41**, 471–484 (1993)
23. Stevenson, S.D.F.: Experimental investigations of lip motion in brass instrument playing. Ph. D. thesis. University of Edinburgh (2009)
24. Kausel, W., Mayer, A., Nachtmann, G.: Experimental demonstration of the effect of wall vibrations on the radiated sound of the horn and a search for possible explanations. In: Proceedings of International Symposium on Musical Acoustics, 2-S.1, Barcellona (2007)
25. Norman, L., Chick, J.P., Campbell, D.M., Myers, A., Gilbert, J.: Player control of 'brassiness' at intermediate dynamic levels in brass instruments. Acta Acust. Acust. **96**, 614–621 (2010)
26. Gilbert, J., Menguy, L., Campbell, D.M.: A simulation tool for brassiness studies. J. Acoust. Soc. Am. **123**, 1854–1857 (2008)
27. Nobara, Y.: Acoustical modelings of the muted brass instruments. Master thesis, Graduate School of Design, Kyushu University (2015)
28. Benade, A.H.: On the mathematical theory of woodwind finger holes. J. Acoust. Soc. Am. **32**, 1591–1608 (1960)
29. Keefe, D.H.: Theory of the single woodwind tone hole. J. Acoust. Soc. Am. **72**, 676–687 (1982)
30. Nederveen, C.J.: Acoustical Aspects of Woodwind Instruments, 2nd edn. Northern Illinois University Press, DeKalb, USA (1998)
31. Lefebvre, A., Scavone, G.P.: Characterization of woodwind instrument toneholes with the finite element method. J. Acoust. Soc. Am. **131**, 3153–3163 (2012)
32. Yoshikawa, S., Kajiwara, K.: Cross fingerings and associated intonation anomaly in the shakuhachi. Acoust. Sci. Tech. **36**, 314–325 (2015)
33. Lighthill, J.: Waves in Fluids, pp. 100–113. Cambridge University Press, New York (1978)
34. Lampton, M.: Transmission matrices in electroacoustics. Acustica **39**, 239–251 (1978)
35. Bertsch, M.: Trumpet mutes. In: 32nd Czech Conference on Acoustics, pp. 1–4. Prague, 23–26 Sept 1995 (1995)
36. Yamada, N.: A study on acoustical effects of the vocal tract on sound generation mechanism of the trumpet. Master thesis, Graduate School of Design, Kyushu University (2010)

Authors Biography

Shigeru Yoshikawa After graduating from Physics Department of Nagoya University in 1974, Shigeru **Yoshikawa** started acoustical research on organ pipes. He worked for Technical R&D Institute of Defense Ministry from 1980 and investigated underwater organ pipes while personally studying musical acoustics. He was a professor of musical instrument acoustics at Graduate School of Design, Kyushu University from 1998 and was retired in 2015.

Yu Nobara After graduating from the Department of Acoustic Design in Kyushu University, Yu **Nobara** went on to the Graduate School of Design in 2013. He made a study on tonal characterization and acoustical modeling of the muted brass instruments. After completing the master course in 2015, he has been working as software designer of car audio systems in Fujitsu Ten Ltd.

Experimental Approaches to the Study of Damping in Musical Instruments

Malte Kob

Abstract Damping plays a central role in the production of musical sounds. Whenever sound sources oscillate, resonators are active or structures radiate sound to the environment, damping is one of the most important parameters in the transmission chain of musical instruments. This contribution shall review the most important damping phenomena that are related to musical instruments and review measurement methods that are employed to quantify damping values. The relations between the various specific values that characterize damping are summarised, and their scopes are discussed.

1 Introduction

Among musical instruments some features can be compared: size, colour, design are visual cues; playing range, ease of playing, projection, timbre and sound intensity are acoustic cues. For the sound of a musical instrument a number of aspects have essential impact on these acoustic cues. The list of these aspects is long, and this one is not complete:

- Material of oscillator such as reed
- Connection between oscillator and resonator
- Construction type of the resonator
- Material of the resonator
- Wave transmission from oscillator to radiator
- Geometry of radiator
- Attachment of the instrument to the player/stand/floor/room.

In all of these aspects conservation, transmission or loss of sound energy can be considered as origin of the perceived sound quality.

M. Kob (✉)
Erich-Thienhaus-Institute, Detmold University of Music,
Neustadt 22, 52756 Detmold, Germany
e-mail: kob@hfm-detmold.de

A. Schneider (ed.), *Studies in Musical Acoustics and Psychoacoustics*,
Current Research in Systematic Musicology 4,
DOI 10.1007/978-3-319-47292-8_6

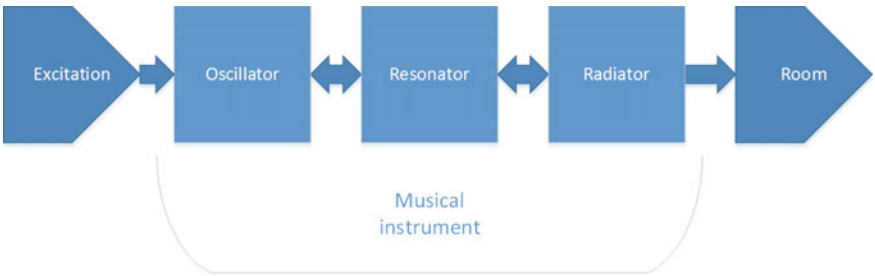


Fig. 1 Scheme of the energy flow in a musical instrument

Figure 1 illustrates the energy flow in a musical instrument. The energy distribution is quite complex in most instruments, and the assumption of a unidirectional transfer from the player who excites the instrument to the listener in the room is only valid as a first-order approximation. In every transition between the segments of the transfer path energy is partially lost from the chain and partially reflected back into the sending segment.

However, many models of music instruments use the source-filter approach as a concept that explains many features of musical sounds such as the generation of formants or radiation from modal structures. This approach is not applicable to more complicated models that shall explain phenomena such as formant tuning or non-linear effects of instruments that exhibit mode-coupling.

Damping plays a major role when a transfer path analysis is done in music instruments; various kinds of damping occur with different effects on the sound and the playability of musical instruments.

For the example of a flute, the self-sustained oscillation of sound waves travelling between the embouchure and the nearest open tone hole is maintained by the player's air flow that feeds energy into the flute for the intended duration. The energy inside the flute is reduced at each instant of the wave reflections at both ends by radiation into the surrounding air, through friction at the boundary of the flute body and through heat diffusion and viscosity in the medium air. More complex effects occur in the damping at the labium through generation of turbulences and vortices that characterize the non-harmonic sound generation in wind instruments [1, 2].

The following section summaries some of the major damping definitions and their definitions.

2 Definition of Damping

Fundamental concepts of damping are reviewed in [3] for oscillating structures and waves in media. In addition to dissipation of energy destructive interference in a standing wave is also considered. An overview of damping mechanisms and a

description of various technical implementations of structural damping is given in Chap. 4 of [4].

An overview of various damping mechanisms and their analogy to electrical circuits is given in [5].

Most literature derives damping from the macroscopic behaviour of an oscillating system such as a mass spring system.

Mass spring system

The classical approach to describe damping of oscillators or waves is derived from the differential equation of an attenuated oscillation of a mass spring system (see Fig. 2).

$$m\ddot{x}(t) + d\dot{x}(t) + cx(t) = F(t). \tag{1}$$

These forces are involved:

$m \cdot \ddot{x}$ inertia force
$d \cdot \dot{x}$ damping force
$c \cdot x$ spring force
$F(t)$ excitation force.

The solution of the homogeneous differential equation of second order is

$$\ddot{x}(t) + \frac{d}{m}\dot{x}(t) + \frac{c}{m}x(t) = 0 \tag{2}$$

written as

$$\ddot{x}(t) + 2\delta\dot{x}(t) + \omega^2 x(t) = 0 \tag{3}$$

with

$\delta = \frac{d}{2m}$: decay constant and

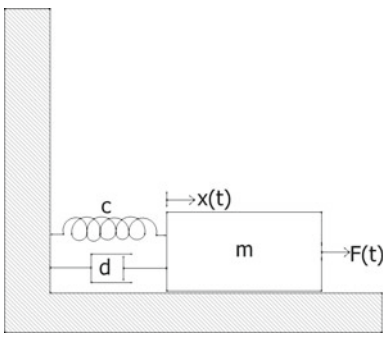


Fig. 2 Mass spring system

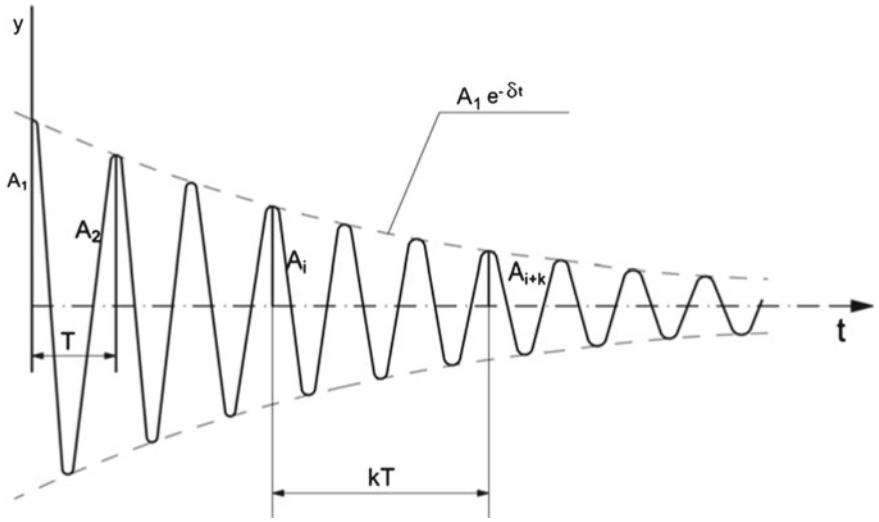


Fig. 3 Damping derived from the decay of an oscillation

$\omega_0 = \sqrt{\frac{c}{m}}$: angular eigenfrequency of the undamped system and the solution

$$\lambda_{1,2} = -\delta \pm \sqrt{\delta^2 - \omega_0^2}$$

Three cases are to be distinguished:

I. $\delta > \omega_0 \Rightarrow \lambda_{1,2} = -\delta \pm \omega_e$: real eigenvalues (sub-aperiodic damping)
II. $\delta < \omega_0 \Rightarrow \lambda_{1,2} = -\delta \pm j\omega_e$: complex eigenvalues (supra-aperiodic damping)
III. $\delta = \omega_0 \Rightarrow \lambda = -\delta$: aperiodic damping (critical damping)

For eq. 1 the time series reads $x(t) = Ae^{-\delta t}\cos(\omega_e t - \Phi)$ and the eigenfrequency reads $f_e = \frac{\omega_e}{2\pi} = \frac{1}{T} = \frac{1}{2\pi}\frac{\sqrt{4mc-d^2}}{2m}$.

In Fig. 3 the damping is indicated as δ for the exponential decay of the oscillation.

3 Classification of Damping

The dissipation of sound and vibration into heat causes damping of amplitudes and a decrease of the quality factor of resonances. This effect is both frequency and space dependent [6].

One approach to classify damping in musical instruments is the categorisation according to the occurrence with respect to the observed sub-system in the transfer path. For the musical instrument this can be done for the excitation, resonator or radiator. For each sub-system an energy entry path, internal energy processing and energy output path can be considered:

- impedance mismatch at entry that reduces energy input from the previous into the observed sub-system,
- energy loss inside the observed sub-system,
- impedance mismatch that reduces the energy flow from the observed into the next sub-system, i.e. conservation of energy in the sub-system.

Figure 4 illustrates this dependency.

In musical instruments, for each of the sub-systems different energy loss mechanisms can be observed which could be a second way to classify damping:

- intrinsic damping due to material properties of the oscillating media,
- damping due to energy flow into the clamping of the medium to surrounding structures,
- damping due to radiation of acoustic or structural waves into air.

These classification approaches are not exhaustive, other categories could be physical definitions or classification according to frequency or amplitude ranges.

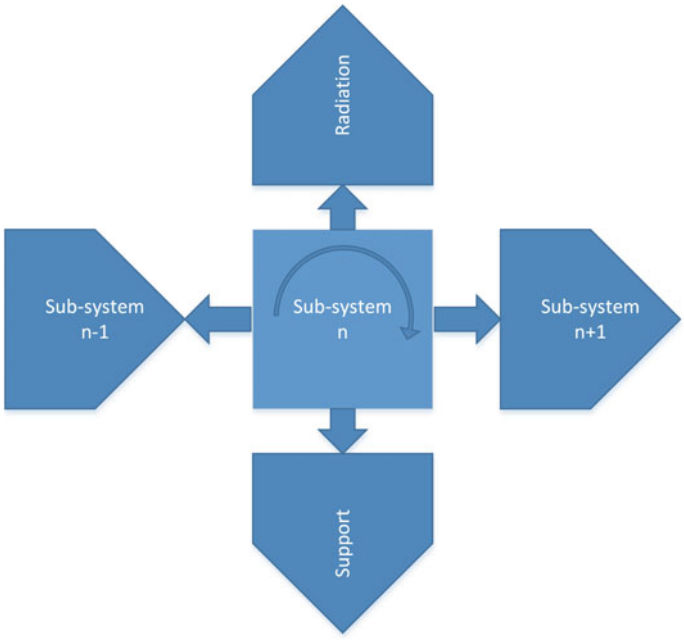


Fig. 4 Energy flow among the sub-systems and into the surrounding

Perceptually, another strong aspect of damping is the transient sound timbre of musical instruments. During build-up and decay of the instrument's sound the order and relative amplitude of harmonic and non-harmonic sound components give musicians and listeners a more or less subtle impression of the sound. Frequency-dependent damping can be a key factor in the perceptual experience [7].

Instrument builders as well as musicians consider the material properties of musical instruments as one of the most relevant factors for their construction. The intrinsic damping of the material would therefore be among the key factors in the design process, together with elasticity, workability, colour, price and others.

4 Measurement Methods

Numerous measurement methods have been established to asses damping values from experimental investigations. Most methods aim at determination of the intrinsic damping of a material. Therefore, under laboratory conditions, the material is shaped, clamped and exposed to oscillations.

Since most materials show a frequency-dependent damping, the evaluation of the measurements is often performed over a large frequency range.

Some of the most frequently used methods are reviewed and compared in this section.

4.1 Torsional Pendulum

Damping can be derived from the assessment of the logarithmic decrement via deflection of a light beam on a torsional pendulum [4]. With the definition of the logarithmic decrement

$$\Lambda = \ln\left(\frac{A_n}{A_{n+1}}\right) \tag{4}$$

as the logarithm of the amplitude ratio of two sequent amplitudes, the loss factor becomes

$$\eta = \frac{\Lambda/\pi}{\sqrt{1+\left(\frac{\Lambda}{2\pi}\right)^2}} \tag{5}$$

and for small damping:

$$\eta = \frac{\Lambda}{\pi} \tag{6}$$

From the oscillation also other parameters such as the shear modulus can be derived.

4.2 Bending Beam

In musical instruments vibrating plates or tongues often play an important role for sound generation or radiation. The model of a bending beam can be used to derive a value for the damping and elasticity. As for the torsion pendulum, an oscillation is provoked, and the decay of the oscillation amplitude is evaluated at various modes (see Fig. 5). Alternatively, several decay values can be derived from at the eigenfrequencies or, alternatively, in frequency bands.

The following variables can be defined for a rectangular beam:

$$f_n = \frac{\pi h a_n^2}{4l^2}\sqrt{\frac{E}{3\rho}}$$

with

$$a_1 = 0.595;\ a_2 = 1.494;\ a_3 = 2.5;\ a_n = n - 0.5;$$

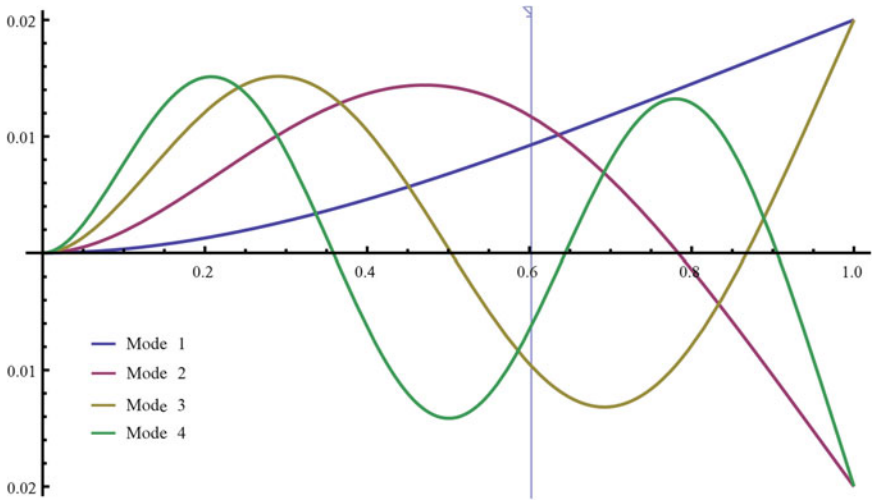


Fig. 5 Modes of a bending beam

h: thickness; l: length; E: Young's modulus; ρ: density resulting in a Young's modulus of

$$E(f) = \frac{48\rho}{\pi^2}\left(\frac{l^2 f_n}{h a_n}\right)^2 \tag{7}$$

and a loss factor of

$$\eta(f) = \frac{6\ln(10)}{T_{60} 2\pi f_n} \tag{8}$$

that is derived from the decay time at a resonance frequency of the beam.

This method has successfully been applied to the investigation of the loss factor in metal organ pipes [8]. Special care was taken to prevent mass loading of the beam using electromagnetic excitation and detection of the oscillating beam.

4.3 Comparison of Different Approaches

Various different measures can be used to describe the damping. Each of them has been derived from a specific method evaluating damping under more or less defined conditions. In Table 1 some of the most frequently used measures and their inter-relation are composed. A more comprehensive table is given in [4].

All methods described above can be applied to samples of instrument materials as well as to whole instruments. Many different implementations of the methods can be applied in the time or frequency domain, and the use of various sensors such as optic, inductive, magnetic or piezoelectric transducers. Finally an impulse or

Table 1 Measures for damping and their inter-relation

Symbol	Name	Definition re. η	Unit
η	Loss factor	$\eta = \eta$	[–]
δ	Decay constant	$\delta = \pi\eta f_n$	[1/s]
Λ	Logarithmic decrement	$\Lambda = \pi\eta$	[–]
T_{60}	Reverberation time	$T_{60} = \frac{6\ln(10)}{2\pi f_n \eta} = \frac{2.2}{f_n \eta}$	[s]
Q	Quality factor	$Q = \frac{1}{\eta}$	[–]
B	−3 dB bandwidth	$B = \eta f_n$	[Hz]
ζ	Damping ratio	$\zeta = \frac{\eta}{2}$	[-]

continuous excitation can be used, and elongation, velocity, acceleration or force can be measured with or without reference to the excitation signal.

This variety of possibilities offers a great spread of more or less useful results of the damping measurements.

From the huge number of analysis methods some few are regularly used to characterize damping in musical instruments.

The evaluation of the reverberation time can easily be used with standard sound level meters and computer programs. Also, the evaluation of the loss factor can be calculated using Eq. 8 which can be simplified further to

$$\eta \approx \frac{2.2}{T_{60} f_n} \tag{9}$$

for an individual frequency. Therefore this method will be characterized further.

A pre-requisite for a frequency-resolved analysis is the application of band filters in room and building acoustics, usually octave or third-octave filters are used.

An interesting aspect of these use cases is the difference in conditions for the various analysis methods. Damping in mass-spring-systems is derived from the decay of a forced vibration, therefore the validity of the damping values is restricted to the eigenvalues of a resonating structure. On the other hand, damping as a material constant should be independant of the resonances of a structure and can be evaluated globally using statistical evaluation approaches at high modal densities such as the reverberation time method. Finally, the reverberation time method could also be applied to single resonances, and the damping values derived from single resonances could be averaged.

The question rises whether all these possibilities yield the same results. Would the evaluation of several resonances within one fractional band filter (as for rooms) be equivalent to the average of single resonances that are separately evaluated? What effect does the application of a narrow-band filter have on the damping value? Is a value obtained from such a method reliable?

These questions shall be answered by a comparison of damping values derived from the two approaches reverberation time evaluation and −3 dB evaluation, applied to the damping measurement of a metal tongue that performs bending motion.

5 Investigations on a Metal Tongue

In the following, two methods for damping assessment are applied to a metal tongue that is clamped on one edge and free on the other sides. This configuration is typical for reed instruments and similar measurements could be performed both on samples of materials or on complete instruments.

The procedures of this chapter and their result have partially been presented at DAGA 2015 [9].

The excitation has been performed both with an impulse hammer and a shaker. The evaluation has been performed using the reverberation time method with impulse hammer excitation in the time domain and the −3 dB method in the frequency domain using shaker excitation and laser Doppler velocimetry.

5.1 Evaluation Using the Reverberation Time Method

The damping has been evaluated using the reverberation time method from the decay of an oscillation of a metal tongue after excitation with an impulse hammer with integrated force transducer. In Fig. 6 the damping is depicted for a combination of hammer excitation and acceleration measurement. The diagram also contains the limit curve of damping that is calculated from the 1/12 octave filters that are used to obtain frequency-resolved damping values. These values are generated from a calculation of the damping via the reverberation times of the band pass filter when excited with a Dirac impulse. This limit curve indicates the maximum damping that any structure may have without being influenced by the equivalent damping of the 12/octave filters used for the evaluation. Consequently, the evaluation of the damping using the reverberation time methods need to

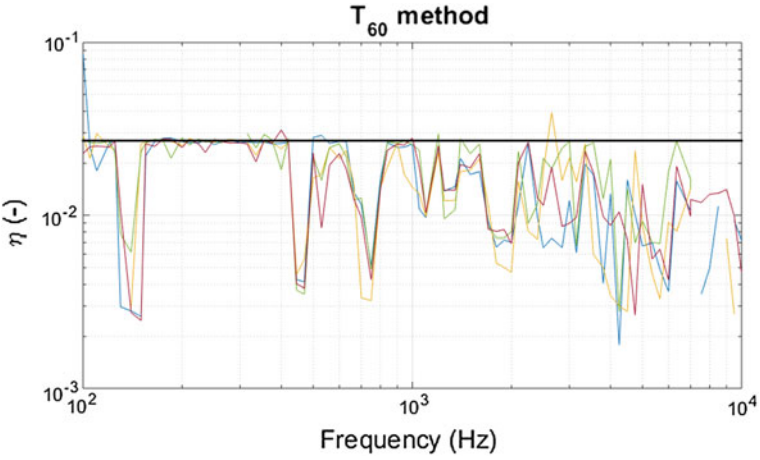


Fig. 6 Damping measurement using an accelerometer and an impulse hammer exciting the surface at several locations. The graph shows the loss factor values evaluated from the reverberation times in 1/12 octave bands together with the loss factor curve of the 1/12 octave filters (*black horizontal line*)

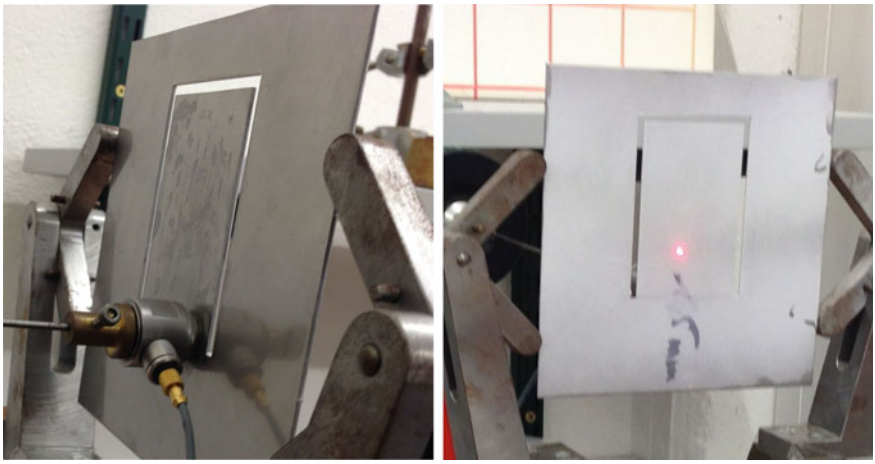


Fig. 7 Exemplary set-up of damping measurement on a metal tongue using laser Doppler velocimetry: excitation with a shaker (*left*) and laser point on the tongue surface (*right*)

consider the increasing filter effect with smaller filter bandwidth respective better frequency resolution [10].

In this analysis the damping would be correctly evaluated whenever the values are smaller than the equivalent filter damping. This occurs at the positions of the resonances-as used in the method of [8]. Outside the resonances the values seem to be not valid because they coincide with the equivalent filter damping.

5.2 Evaluation Using the −3 dB Method

In Fig. 7 the set-up for measurement of the damping of the metal tongue is shown using a shaker and laser Doppler velocimetry. A excitation point an asymmetric location at the lower end of the tongue was selected. The laser scanned a mesh of 200 points.

The frequency transfer function is represented as real part of the mobility $\frac{v}{N}$ versus frequency (see Fig. 8). The damping is evaluated from the frequency values at −3 dB on both sides of each resonance. At resonances with high quality (e.g. at 150 or 1500 Hz) the damping—that otherwise falls monotonously with rising frequency—has significantly smaller values. The evaluation was performed automatically with a MATLAB script.

Whereas the evaluation of high damping values from reverberation times using fractional octave filters is limited, no such effect is expected here. Therefore all calculated values should be valid. Uncertainties can arise from errors of the bandwidth evaluation in case of neighboured or weak resonances.

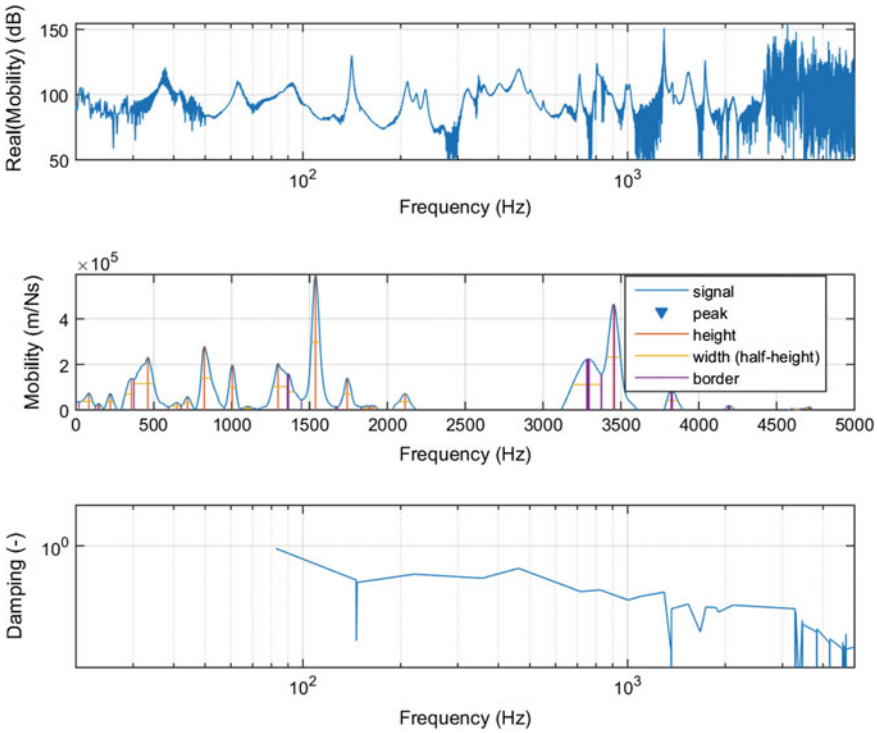


Fig. 8 Example of −3 dB method using shaker and laser

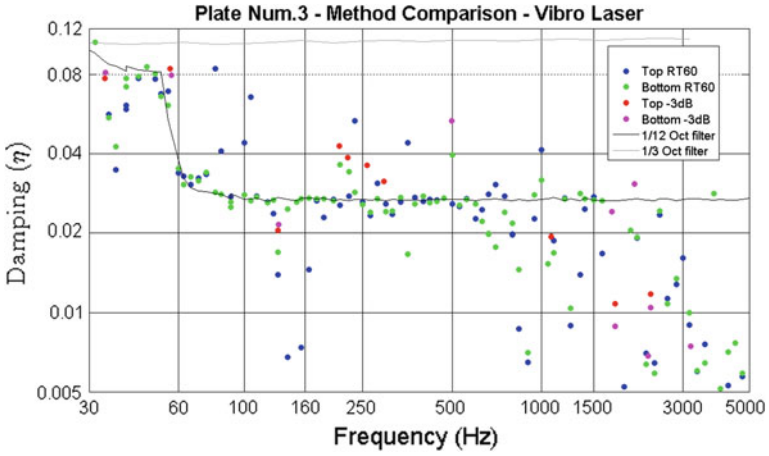


Fig. 9 Evaluation of two methods, together with limit curves of 1/3 and 1/12 octave filters (*black and gray*)

5.3 Interpretation and Comparison of Both Methods

A comparison of the two methods indicates a strong variation of the damping results in dependence of the location of the measurement and the choice of the method.

In Fig. 9 the limit curves are depicted for 1/3 and 1/12 octave filters, together with the damping values derived from different evaluation methods.

The excitation was realised with an impulse hammer, and the transfer function $\frac{v}{N}$ was calculated. The functions were evaluated with excitation on top of the plate and laser measurement on top or at the bottom. The damping values derived from the −3 dB method measured at the top, free end of the plate are represented with red dots, the values measured at the bottom, near the clamped edge, are shown in violet. The corresponding values evaluated with the reverberation time method are shown in blue (top) and green (bottom).

A correlation of the values with the position of the measurement can be observed, but not for all combinations of frequencies and methods.

Interesting is the observation of higher damping values around 250 Hz that would fall in the limit curve of 1/12 octave filters when using the reverberation time method. A less selective evaluation using 1/3 octave filters would not have that limitation.

The evaluation of the vibration at the upper end of the metal tongue seems to yield lower damping values at low and medium frequencies when using the RT60 method. For this sample, both methods seem to indicate a dependency of the loss factor with frequency. However, this dependency does not seem to be linear: high values at low frequencies, more or less constant values at mid frequencies (with exception of the resonance peaks) and decreasing values above ca. 1 kHz.

6 Conclusion

Several methods for evaluation of damping in musical instruments can be used. The results from the analysis methods, however, differ with respect to the choice of excitation and measurement point, filtering method and resonance conditions.

The −3 dB method is more demanding with respect to the evaluation at single resonances but does not require a filter that can limit the validity of the results. At higher frequencies both methods seem to give more spread but unlimited (by filter effect) values.

Generally lower damping values are observed at resonances with high quality. This could be an approach to distinguish intrinsic and structural damping. More investigations are needed to confirm this assumption.

Alternative methods for evaluation of damping could be designed during playing of the instrument from the evaluation of the excited resonances of the enclosure of

the instrument using high speed camera images [11]. As of today these methods do not yet offer the required frequency resolution and dynamics.

Acknowledgments Albrecht Schneider is acknowledged for fruitful discussions and providing literature sources. Dustin Eddy is acknowledged for the assistance in the evaluation of the measurement results. The author is grateful to Michael Vorländer and Markus Müller-Trapet for the opportunity and support during the laser vibrometer measurements at the Institute of Technical Acoustics at RWTH Aachen University.

References

1. Fabre, B.: La production de son dans les instruments à embouchure de flûte: modèle aéro-acoustique pour la simulation temporelle. PhD thesis, Université du Maine, Le Mans (1992)
2. Hirschberg, A.: Mechanics of Musical Instruments, Chapter Aero-Acoustics of Wind Instruments. Springer, New York (1995)
3. Wagner, K.W.: Einführung in die Lehre von den Schwingungen und Wellen. Dieterisch'sche Verlagsbuchhandlung (1947)
4. Cremer, L., Heckl Björn, M., Petersson, A.T.: Structure-Borne Sound, 3rd edn. Springer, Berlin (2005)
5. Meyer, E., Guicking, D.: Schwingungslehre. Vieweg (1974)
6. Chaigne, A., Kergomard, J.: Acoustic of Musical Instruments. Springer, New York (2016)
7. Castellengo, M.: Ecoute musicale et acoustique: Avec 420 sons et leurs sonagrammes décryptés. Eyrolles (2015)
8. Svensson, P., Friesel, M.: Influence of alloy composition, mechanical treatment and casting technique on loss factor and young's modulus of lead-tin alloys. Technical report, Chalmers University of Technology (1999)
9. Kob, M., Eddy, D.: Evaluation verschiedener experimenteller Ansätze zur Charakterisierung der Materialeigenschaften von Musikinstrumenten. In: Fortschritte der Akustik - DAGA 2015 (2015)
10. Kob, Malte, Vorländer, Michael: Band filters and short reverberation times. Acustica **86**(2), 350–357 (2000)
11. Grothe, Timo, Barth, Rainer, Kob, Malte, Jaschinski, Jörn: Hochgeschwindigkeits-Stereophotogrammetrie eines klingenden Stabes. Fortschritte der Akustik-DAGA **2015**, 818–821 (2015)

Author Biography

Malte Kob was born in Hamburg where he was educated as church musician. After studies of electrical engineering and a diploma thesis with the Physikalisch-Technische Bundesanstalt (PTB) in Brauschweig, he focussed on voice acoustics in his dissertation approved by the RWTH Aachen University. Since 2009 he coordinates the engineering education and music acoustics programs as Professor at the Tonmeister Institute of the Detmold University of Music.

Comparison of Vocal and Violin Vibrato with Relationship to the Source/Filter Model

James W. Beauchamp

Abstract Although the source/filter model is often mentioned in the literature of acoustics and signal processing (e.g., Gold and Morgan, *Speech and Audio Signal Processing*, Wiley), it has seldom been implemented for musical instrument sounds. For operatic style male voices with sufficient vibrato depth, overlapped harmonic amplitude-versus-frequency (HAF) graphs can yield displays that trace out vocal tract resonances quite effectively (Maher and Beauchamp in Appl Acoust 30:219–245, 1992 [4]; Arroabarren and Carlosena J Acoust Soc Am 119(4):2483–2497, 2006 [5]). If the glottus signal can be derived, its spectrum (in dB) can be subtracted from the HAF data to reveal a vocal tract transfer function. However, for the violin the HAF method with vibrato excitation has proved unsuccessful because (1) violin vibrato depths are generally insufficient and (2) HAF traces appear too steep to be caused by actual violin resonances. Therefore, a violin glide tone (C5-to-C4, performed in an anechoic chamber) was used instead. Based on an assumption that the source signal spectrum was independent of fundamental frequency (F0), average ratios between adjacent harmonic amplitudes were measured making it possible to derive a source spectrum (within a scale factor). From this the violin transfer function was derived. As a comparison, a pair of violin glide signals (one at the bridge and the other radiated) supplied to the author by Alfonso Perez-Carrillo (J Acoust Soc Am 130(2): 1020–1027, 2011 [19]) were analyzed. The measured bridge spectrum was similar to that of the C5-to-C4 tone's derived source spectrum, but the derived filter was quite different, as might be expected considering the different violins and arbitrary microphone positions used in the two cases.

J.W. Beauchamp (✉)
School of Music and Department of Electrical and Computer Engineering,
University of Illinois at Urbana-Champaign, 1002 Eliot Dr, Urbana, IL 61801, USA
e-mail: jwbeauch@illinois.edu
URL: http://ems.music.uiuc.edu/beaucham

A. Schneider (ed.), *Studies in Musical Acoustics and Psychoacoustics*,
Current Research in Systematic Musicology 4,
DOI 10.1007/978-3-319-47292-8_7

1 Introduction

In the 1930s Seashore [1] published a book which included an extensive study of vibrato. Since then many studies have been completed, notably those by Sundberg [2, 3]. Using a drawing, Sundberg showed a relationship between harmonic frequency and amplitude vibrato based on a vocal formant characteristic [2]. This was verified by Maher and Beauchamp [4] using a time-varying spectrum analysis to graph the amplitude-versus-frequency characteristics of vocal sounds. With a vibrato depth approaching ±6 % (±half-step), the first few harmonic traces are distinct, but starting at harmonic n where $nf_1(1 + 0.06) \geq (n + 1)f_1(1 - 0.06)$ (so in this case harmonic $n = 8$) the harmonics begin to overlap and trace out a cohesive formant characteristic. More recently Arroabarren and Carlosena [5] showed that the amplitude-versus-frequency representation provides a spectral envelope very similar to a corresponding vocal tract response characteristic obtained by inverse filtering.

Maher and Beauchamp [4] also showed how the peak magnitude of a vocal vibrato's FFT spectrum can be used to give a precise indication of the average vibrato rate. The longer the vibrato's sinusoidal behavior lasts, the more the FFT's peak approaches a vertical line. The authors also presented data for parameters describing vocal vibrato waveforms, including random variations of vibrato mean, depth, and rate.

Although not as common as those for the singing voice, a few relevant studies of violin vibrato have been produced. Fletcher and Sanders [6] reported that whereas percent frequency depths of vibrato were the same for all harmonics measured, there was a considerable difference between the amplitude variations of the harmonics. Detailed curves of frequency vibrato and amplitudes of the first 10 harmonics were shown, but unfortunately the method by which these curves were derived was not elucidated. Mellody and Wakefield [7] used a high-resolution time-frequency method that derived amplitude as well as frequency modulations of the harmonics. Only certain harmonic amplitudes exhibited vibrato-like oscillations in a traditional violin, but these oscillations did not appear in a solid body violin. In the first case the underlying "note evolution" (presumedly due to bowing) and the more rapidly varying "residual" vibrato characteristic could be separated. Both of these studies incorporated bow noise for synthesis of violin tones. Bow noise was found to be important for 100 % discrimination of synthetic from original tones but for judgement of similarity in a multidimensional scaling study only for high F0's ($\geq$ 550 Hz). This qualitatively agrees with the Fletcher and Sanders [6] study which noted that bow noise was hardly noticeable for F0 < 400 Hz.

Since predicting the amplitude modulation of vocal harmonics based on a model for the source waveform and corresponding spectrum together with a filter function proved to be quite successful, it seems that a source/filter model would be useful in the case of the violin. Like the vocal cords and the vocal tract, the vibrations of the

string and the multi-resonant characteristic of the body appear to be quite independent. However, measurements of source spectra or filter responses for the violin are quite uncommon. Shortly after the necessary electronic technology became available in the 1930s Saunders [8] made "response curves" of several Stradivarius violins by performing a series of pitches (by "hand bowing") in an acoustic environment with very low echo while keeping the bow position, pressure, velocity, and violin orientation as constant as possible with respect to a microphone which fed the signal to a spectrum analyzer that could plot individual harmonics versus frequency. However, he didn't attempt to vertically align the note spectra but rather simply superimposed and averaged them. In the early 1970s Hutchins [9] was able to plot the frequency response of violins over a 20–20,000 Hz range using a *moving armature* mounted on the violins' bridge driven by an audio oscillator. Meinel showed similar graphs [10], based on his work in the 1930s. In the low frequency region the violin's response is dominated by an *air resonance* and a *wood resonance*. Above these resonances the response is very complex. For synthesis purposes Mathews and Kohut [11] approximated the response with 37 resonances.

The source waveform/spectrum has been studied separately in several studies. Using a magnetic-field-induced voltage across a hand-bowed metallic violin string, Kohut and Mathews [12] in 1971 recorded string velocities at several positions along a violin string and found that string motion under ordinary bowing conditions (proper force and velocity) closely approximates Helmholtz motion. However, departures were noted in terms of rise time and superimposed ripple. In 1973 Schelleng [13] examined how a quintessential bowed string displacement waveform (the Helmholtz motion) could be derived for a very flexible string; it results in a sawtooth displacement wave at a very stiff bridge. He then compared the spectrum of a sawtooth, with its $1/n$ harmonic amplitudes, with the spectrum of a real stiff string whose spectrum rolls off more rapidly, especially beyond a certain harmonic. However, depending on playing style, other more complex waveforms are possible with the bowed string. In fact, Raman [14] studied many of these waveforms much earlier in the 20th century, by both measurement and mathematical theory. Schumacher [15] derived an integral formula and numerically calculated the velocity waveform at a certain bowing point. McIntyre and Woodhouse [16] derived a similar result published in the same issue of Acustica. Lawergren [17, 18] went further for *S-motion* waveforms on a bowed string to derive a closed-form equation which matched waveforms plotted from a special apparatus consisting of a bowing machine, a magnetic pickup similar to one used by Kohut and Mathews [12], and a chart recorder. S-motion waveforms, which are superimposed on the Helmholtz waveform, were found to vary drastically depending on bow velocity and relative values of bow position and position of observation. However, these waveforms, which contain a substantial amount of ripple, can only be obtained with a bow force of "operationally defined critical value" and perhaps only by bowing within certain regions of the string.

Deriving a source/filter model of a musical instrument presupposes that both the source waveform or spectrum and the filter transfer function can be measured, either by physical measurement, analytic calculation, computer simulation, or using signal processing methods. However, another approach is to only measure the filter transfer function and use a solid body violin to generate the string signal. This is the approach Mathews and Kohut [11] used except that their filter consisted a number (24–37) of first-order resonance circuits, which were combined in parallel and whose frequencies and bandwidths, rather than being chosen to match the characteristic of a particular violin, were adjusted to produce a pleasing violin-like sound. More recently Pérez Carrillo et al. [19] measured both the time-varying source spectrum at the bridge of a violin using a commercial bridge pickup and the time-varying output spectrum (in various locations) measured in an anechoic chamber to calculate the transfer function and its corresponding impulse response. Again the objective was to enable connection of a solid body violin to a real-time transfer function so as to produce a realistic acoustic violin sound.

2 Vocal Vibrato Analysis

Several operatic-style vocal sounds [20] with significant vibrato depth have been analyzed using a method in the SNDAN musical sound analysis package [21, 22] which is an implementation of the McAulay-Quatieri (MQ) algorithm [23]. Frequency-versus-time vibrato signals are quasi-sinusoidal with significant departures from a pure sinusoidal form. This is based on a weighted average of the first five harmonics where the weights are based on their relative amplitudes. Thus, the *weighted-average frequency deviation* $\Delta f_0(t)$ around the mean fundamental frequency f_0 is given by

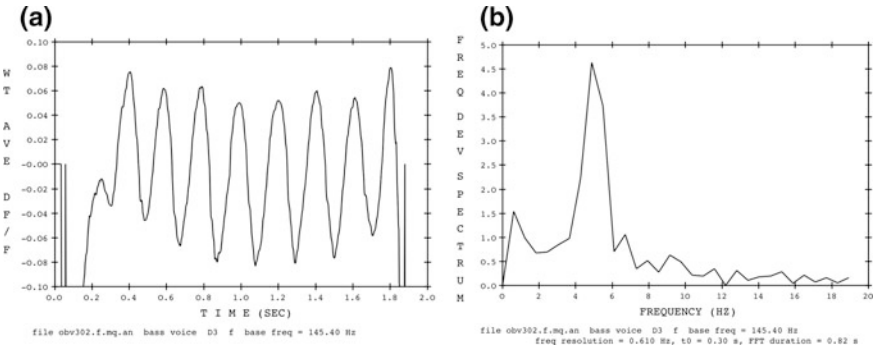


Fig. 1 Graph of the frequency vibrato signal (normalized frequency deviation) for a bass voice in fractional units around a mean fundamental frequency of f_0 = 145.4 Hz (≈D_3): **a** time domain; **b** frequency domain. The maximum value corresponds to a vibrato frequency of 4.9 Hz

$$\Delta f_0(t) = \frac{\sum_{k=1}^{5} A_k(t)\Delta f_k(t)}{\sum_{k=1}^{5} A_k(t)} \tag{1}$$

where $A_k(t)$ and $\Delta f_k(t)$ are the time-varying amplitude and frequency deviation of harmonic k . Then, the *normalized frequency deviation* is the fractional change of f_0, given by $\Delta f_{\text{norm}}(t) = \Delta f_0(t)/f_0$. Figure 1a is a plot of $\Delta f_{\text{norm}}(t)$ for a bass voice signal pitched at D_3, and Fig. 1b shows the magnitude of its discrete Fourier transform.

Note that the total frequency for each harmonic k is given by $f_k(t) = kf_0 + k\Delta f_0(t)$. It is assumed that the frequencies of all partials are integer multiples of a common fundamental.

Figure 2a gives a composite spectrogram showing all harmonics varying with respect to frequency in the range $0 < f < 4000$ Hz, where amplitude is indicated by relative darkness. This includes 27 harmonics based on f_0. Figure 2b gives the same

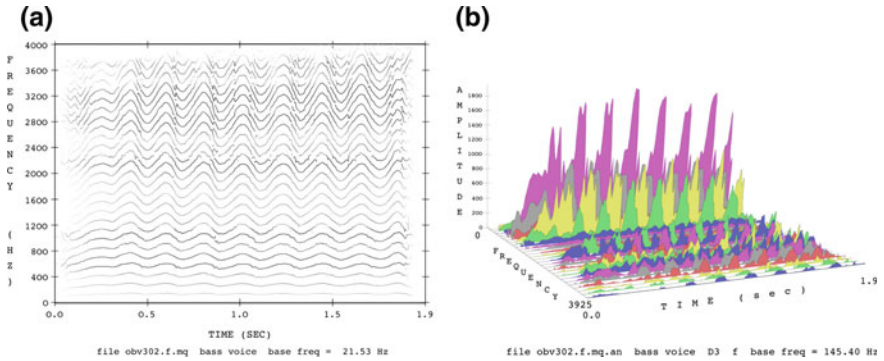


Fig. 2 Spectrograms showing **a** frequency variations; **b** amplitude variations (linear scale) of a D_3 bass voice sound's harmonics

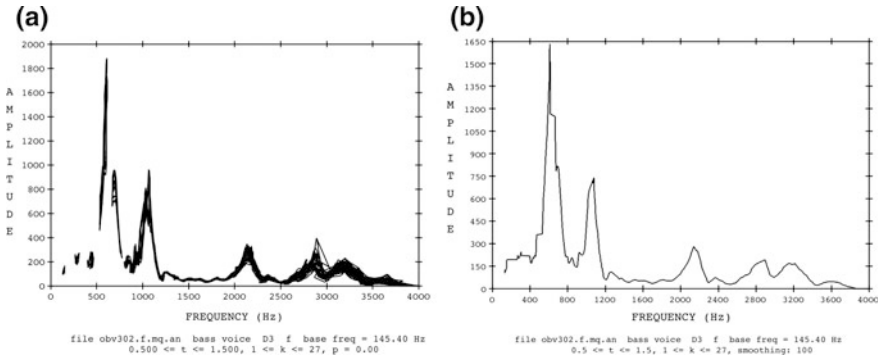


Fig. 3 Spectral envelopes (linear amplitude scale) for bass voice sound: **a** result of overlaying harmonic amplitude-versus-harmonic frequency plots. **b** Smoothed version of (**a**)

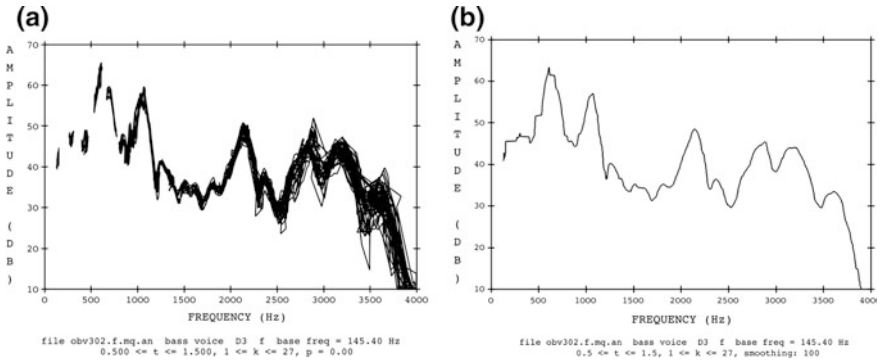


Fig. 4 Spectral envelopes (decibel amplitude scale) for bass voice sound: **a** Result of overlaying harmonic amplitude-versus-harmonic frequency plots. **b** Smoothed version of (**a**)

data as a three dimensional graph showing the amplitudes of the harmonics with line hiding. The vertical axis gives linear amplitude.

In order to show the overall spectral envelope of the vocal sound, we first overlay graphs of $A_k(t)$ versus $f_k(t)$ as shown in Fig. 3a. Then in Fig. 3b, after sorting the Fig. 3a data in frequency order, we use a moving average filter to achieve a smooth spectral envelope. These spectral envelopes are also plotted using decibel amplitudes as shown in Fig. 4.

The curve of Fig. 3b (or Fig. 4b), which we can refer to as the transfer function $H(f)$, can then be driven by a flat spectrum signal (i.e., delete s(t)) with the vibrato pattern given in Fig. 1a to recreate an approximation to the original time-varying spectrum. This constitutes a source/filter model with a unity source spectrum. In terms of a mathematical expression the output of the model could be

$$s(t) = \sum_{k=1}^{K} H(kf_0 + k\Delta f_0(t)) \cos\left(kf_0 t + \int_0^t k\Delta f_0(t)dt + \theta_{0_k} \right), \tag{2}$$

However, the source spectrum, which is all 1's in this case, wouldn't correspond to the spectrum of a real physical source, the spectrum of a glottis waveform. Arroabarren and Carlosena [5] discuss methods for determining source spectra from established parameters of the glottis waveform.

The vibrato waveform signal $\Delta f_0(t)$ can be further parameterized. We propose the following form:

$$\Delta f_0(t) = f_a(t) + d_v(t)\cos(2\pi f_v t + \phi_0), \tag{3}$$

where $f_a(t)$ is an additive frequency drift and $d_v(t)$ is a time-varying vibrato depth. Figures 5 and 6 show estimated curves for $f_a(t)$ and $d_v(t)$ for the D_3 bass voice tone

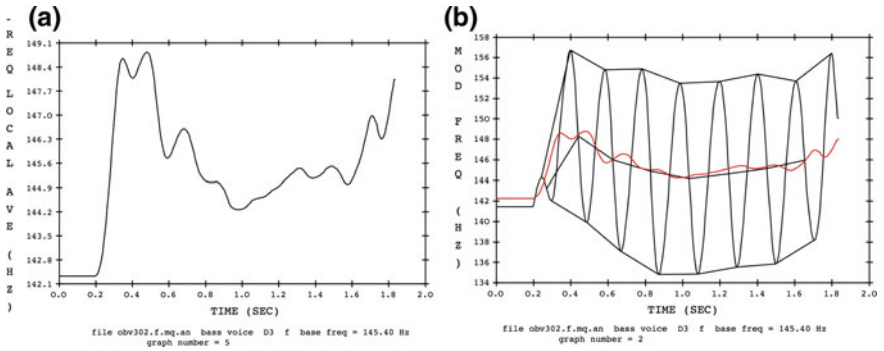


Fig. 5 Frequency drift **a** $f_a(t)$ alone. **b** $f_a(t)$ (in *red*) superimposed on the smoothed vibrato waveform; *upper line* segments connect vibrato peaks; *lower line* segments connect vibrato troughs; *middle line* connects averages of the peaks and troughs

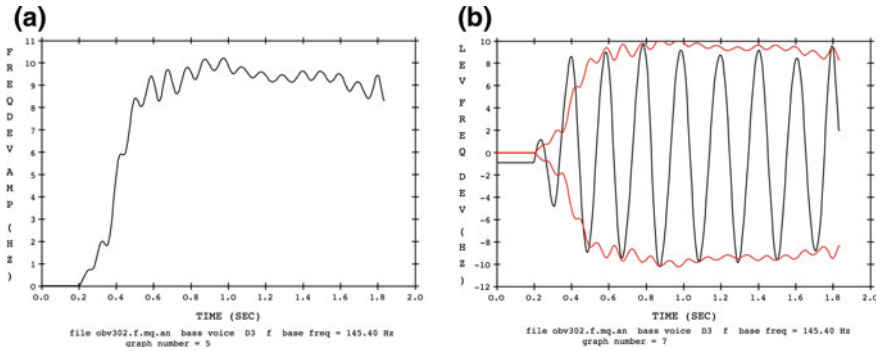


Fig. 6 Time-varying vibrato depth **a** $d_v(t)$ alone. **b** $\pm\, d_v(t)$ (in *red*) superimposed on the leveled vibrato waveform $\Delta f_0'(t)$, which has been slightly modified at its start and finish

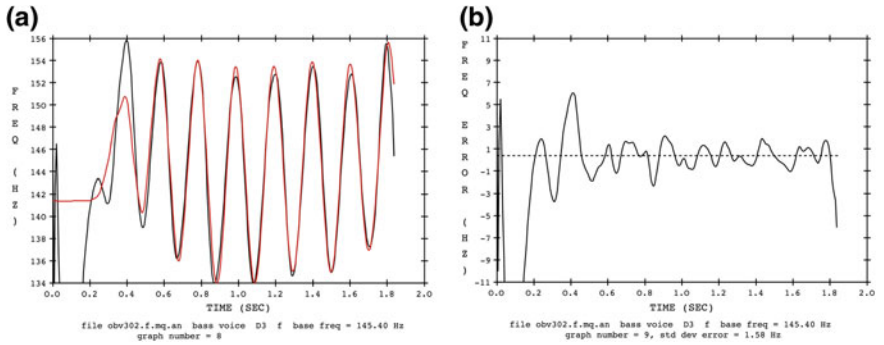


Fig. 7 **a** Reconstructed vibrato waveform according to Eq. 3 (in *red*) overlaid on original smoothed vibrato waveform. **b** Error between original and original vibrato waveforms (std error ≅ 1.6 Hz.)

whose vibrato waveform is shown in Fig. 1a. A moving average with window size $1/f_0$ (where f_0 = 4.88 Hz) was used for $f_a(t)$, and a heterodyne/filter method was used to determine $d_v(t)$. The latter involves multiplication of the "leveled" vibrato waveform (i.e., $f_0'(t) = f_0(t) - f_a(t)$) by $\cos(2\pi f_v t)$ and $\sin(2\pi f_v t)$, averaging over $1/f_0$, and taking the square root of the sum of squares to obtain the $d_v(t)$ estimate. Figure 7a shows an overlay of the original vibrato waveform and the waveform reconstructed from Eq. 3 with the parameter estimates of Figs. 5 and 6. With more perfect parameter estimates, the error could be reduced further, but for vibrato synthesis these estimates are probably adequate.

Equation 2, which gives an equation for the time-varying vocal spectrum, does not include the effect of the source spectrum, which we may assume is independent of f_0 and independent of the vocal tract characteristic, which depends on the vowel sung. This assumption modifies Eq. 2 as follows:

$$s(t) = \sum_{k=1}^{K} A_k H(kf_0 + k\Delta f_0(t)) \cos\left(kf_0 t + \int_0^t k\Delta f_0(t)dt + \theta_{0_k} \right), \tag{4}$$

In this case each harmonic's amplitude-versus-frequency curve that overlaps with neighboring harmonics in frequency will not strictly overlap with its neighbors in amplitude, depending on the ratios of the source harmonic amplitudes. As can be seen from Figs. 3a and 4a, overlap begins at about the 6th harmonic. This can be predicted from the vibrato waveform. We see from Fig. 1a that the extent (aka depth) of the vibrato is roughly ± 8.0 %, which means the frequency range of harmonic k is $(0.92kf_0, 1.08kf_0)$. Frequency overlap between harmonic k_1 and k_2 will occur if $1.08k_1 > 0.92k_2$. For adjacent harmonics where $k_2 = k_1 + 1$, this leads

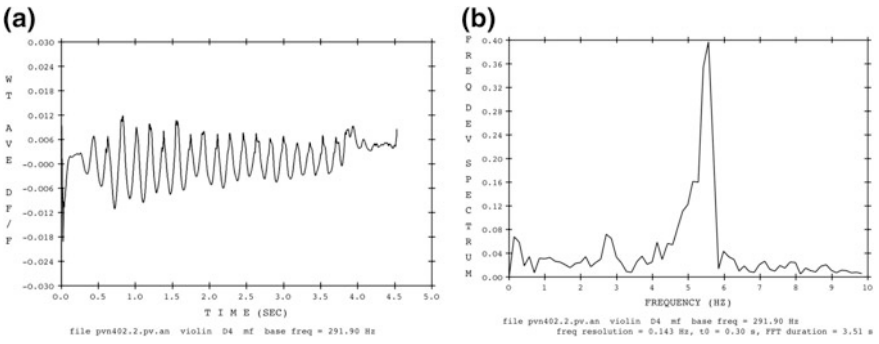


Fig. 8 Graph of the frequency vibrato signal (normalized frequency deviation) for a violin tone in fractional units around a mean fundamental frequency of f_0 = 291.9 Hz (≈D_4): **a** time domain; **b** frequency domain: the maximum value corresponds to a vibrato frequency of 5.6 Hz

to $k_1 > 5.75$ or $k_1 = 6$. (In general, if d is the fractional deviation, this leads to $k_1 > (1 - d)/(2d)$ for overlap to occur.)

According to the assumption that A_k depends only on harmonic number but not on time or frequency, if adjacent harmonics k and $k + 1$ have the same frequency f (albeit at different times), then the ratio of the amplitudes at the input is $r_k = A_{k+1}/A_k$ and the ratio of the amplitudes at the output is $A_{k+1}\, H(f)/A_k\, H(f) = A_{k+1}/A_k = r_k$. In other words, the ratio of adjacent harmonic amplitudes at the output is the same as at the input. Following this reasoning, if we take $A_1 = 1$, we have $A_2 = r_1$, $A_3 = r_1 r_2$, $A_4 = r_1 r_2 r_3$, etc., which then gives us a valid identification of the source spectrum. Then, if we divide the output spectrum, which is $\{A_k H(kf_0)\}$ in our model, term by term by $\{A_k\}$, we will have an improved estimate of the filter response $H(f)$.

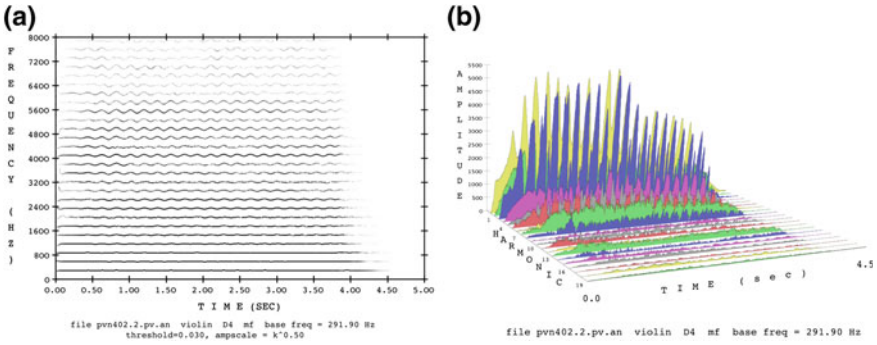


Fig. 9 Spectrograms showing **a** frequency variations; **b** amplitude variations of the D_4 violin tone's harmonics (linear amplitude scale)

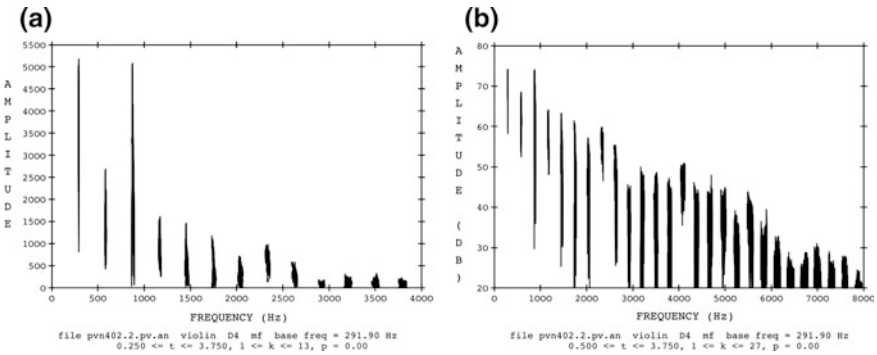


Fig. 10 Amplitude-versus-frequency plots for a D_4 violin tone: **a** linear amplitude scale; **b** decibel amplitude scale

3 Violin Vibrato Analysis

The situation for analysis of a possible source/filter model of the violin would seem similar to that of the voice because, like the voice, the string vibration is largely independent of the violin body resonances which serve as the filter. The ideal force signal at the bridge, the point of input to the violin body, is a sawtooth waveform, which is well known to have harmonic amplitudes proportional to $1/k$. However, Schelleng [13] has shown that for stiff strings typical of the violin, harmonic amplitudes drop off more rapidly than $1/k$.

Using violin vibrato tones to predict the separation between source and filter is more problematic. For one thing, typical violin vibrato depths are much less than for the operatic singing voice. Second, violin resonances appear to be much sharper than those of the voice. So for these reasons the harmonic amplitude-versus-frequency traces do not tend to overlap. Figure 8 shows the vibrato waveform and its Fourier magnitude spectrum for a D_4 violin tone. Figure 9a shows a 2-dimensional spectrogram of the violin, and Fig. 9b shows a 3D graph of amplitude-versus-harmonic-versus-time. Figure 10a, b show amplitude-versus-frequency graphs for the violin tone, and 10b shows that no harmonics overlap up through harmonic 27. Since Fig. 8a indicates that the maximum fractional frequency deviation is about $d = 0.012$, harmonics will not begin to overlap until $k = 41$. Even so, the curves in Fig. 10a, b are remarkable because they are practically vertical. Unlike amplitude-versus-frequency plots for the voice, they don't seem to trace out a recognizable resonance characteristic. Rather it appears that each harmonic traces out its own narrow-band resonance.

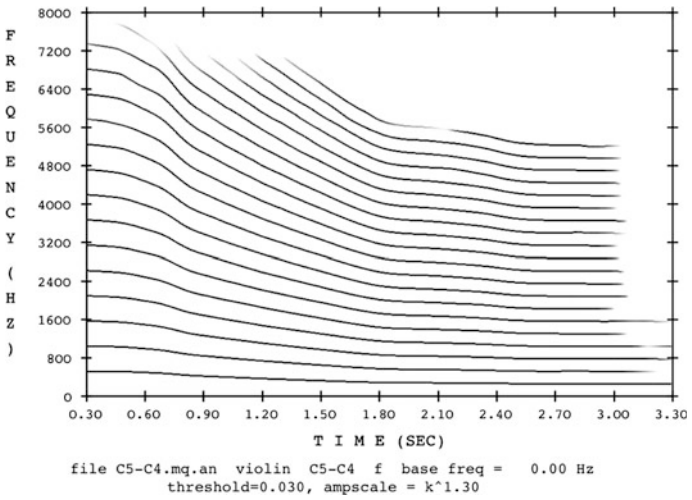


Fig. 11 Harmonics of a violin glide tone starting at C5 (523 Hz) and ending at C4 (262 Hz)

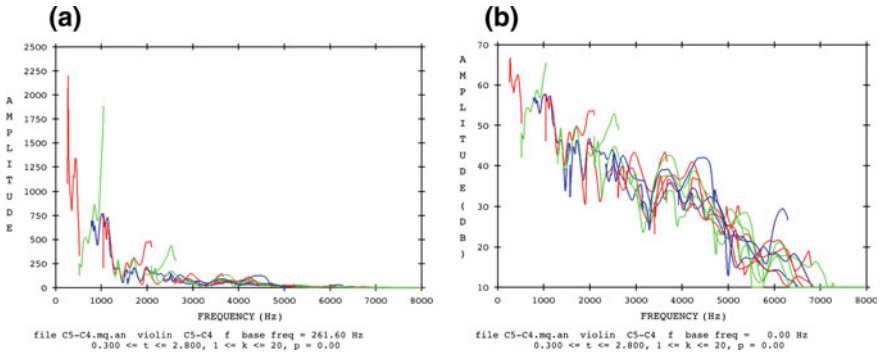


Fig. 12 Overlays of violin harmonic amplitude-versus-frequency tracks for harmonics 1–20 for a C5-C4 glide tone (recorded in an anechoic chamber): **a** linear amplitude; **b** decibel amplitude

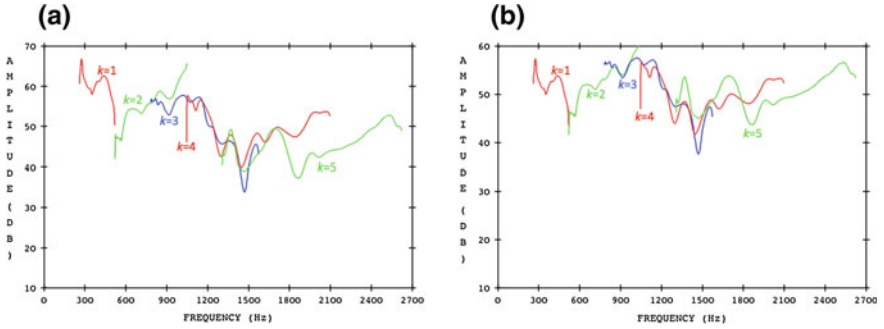


Fig. 13 Overlays of violin harmonic amplitude-versus-frequency tracks for harmonics 1–5 for the C5-C4 glide tone: **a** original; **b** after vertical alignment

4 Violin Frequency Response: The Filter Characteristic and Source Spectrum

The filter characteristic for a violin can be estimated if the output harmonics can be made to overlap on an amplitude-versus-frequency plot, under the assumption that the input harmonic amplitudes are independent of fundamental frequency. To enable this idea, a frequency glide was performed in an anechoic chamber at the Indian Institute of Science, Bangalore, India [24]. Even though the pitch change was just an octave (C5 to C4), this glide resulted in a frequency response over several octaves because 20 harmonics contributed to the response. The glide tone's harmonic frequencies are shown in the 2-dimensional spectrogram graph in Fig. 11.

The next step is to frequency-track the violin tone harmonics with the MQ method [23] and separate them using a pitch detector and harmonic separator

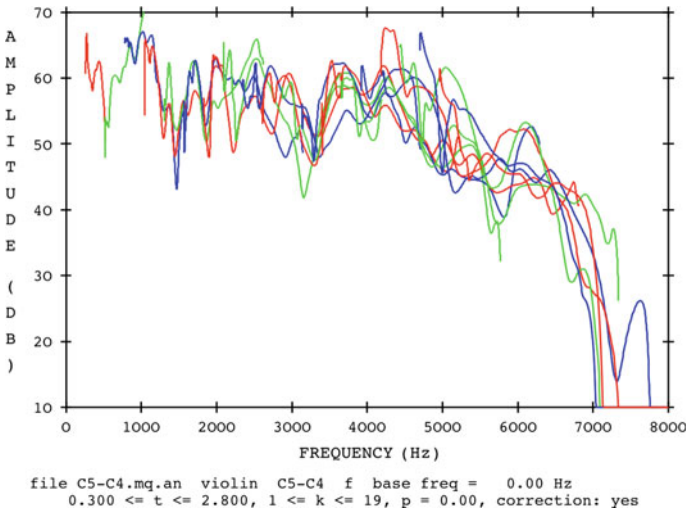


Fig. 14 Overlays of the violin C5-C4 glide tone's harmonic amplitude-versus-frequency tracks for harmonics 1–19 after harmonics were vertically aligned to minimize the error between adjacent harmonics. This may be compared to Fig. 12b, where the harmonics are not vertically aligned

[22, 25]. Figure 12 shows the amplitude-versus-frequency overlaid harmonic amplitude analysis for the violin glide tone, for both the linear and decibel amplitude cases. Here, no effort is made to align the harmonics that overlap, which include all of the harmonics except for the first and the second. (Plots similar to this were made by Saunders [8], except that he used several single tones at different pitches and made no attempt to vertically align the harmonic tracks.) Zooming in on the first 5 harmonics, as shown in Fig. 13a, we can see that there is a fairly good correlation between their amplitude-versus-frequency traces. However, this is improved in Fig. 13b by vertically shifting the individual harmonic tracks relative to one another (i.e., an alignment) to minimize the average decibel error between them.

Note that this vertical alignment in decibels is just a matter of adding an optimum constant positive decibel value to each track, which in linear amplitude units is equivalent to multiplying each kth harmonic track by the inverse of the product $r_1 \ldots r_k$, as defined above. The decibel value added to each track is then

$$dB_k = 20\log_{10}\left(1/\prod_{n=1}^{k=1} r_n\right) = \sum_{n=1}^{k=1} 20\log_{10}(1/r_n) \tag{5}$$

Without loss of generality we assume that the source signal's first harmonic amplitude is unity ($A_1 = 1$). Since in our case, because the glide is only one octave,

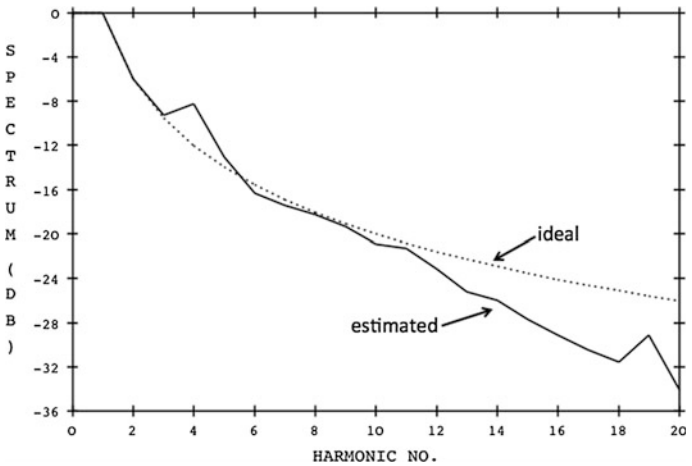


Fig. 15 Estimated source spectrum dB_k for the violin based on a C5-C4 glide tone. The dotted curve is the ideal spectrum given by 20 $\log_{10}(1/k)$, where k is harmonic number

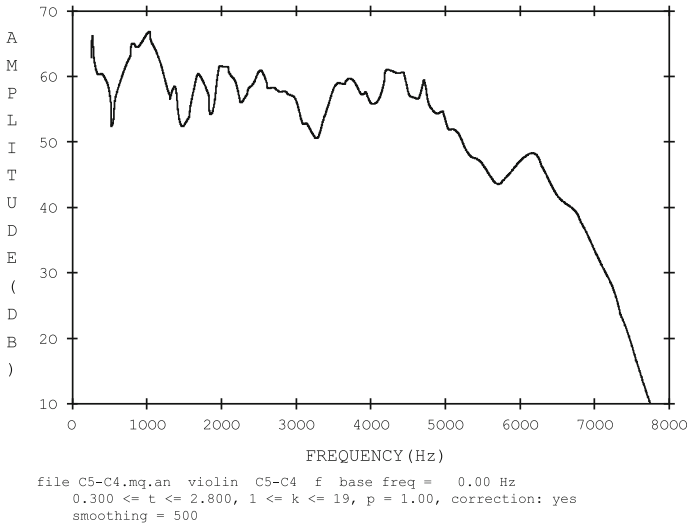


Fig. 16 Estimated filter characteristic for the violin based on the C5-C4 glide tone. This curve was derived by smoothing the data, consisting of 19 harmonic tracks, shown in Fig. 14

the first and second harmonics don't overlap in frequency, and so we are forced to make the reasonable assumption that $A_2 = r_1 = 0.5$. Then if $r_2 = 2/3$, $A_3 = 1/3$ and so forth. To achieve the case $A_k = 1/k$, we need $r_k = k/(k + 1)$. However, this is a

conjecture. Suppose instead that $A_3/A_2 = r_2 = 0.6$. Then, we would have $A_3 = 0.5 \times 0.6 = 0.3$.

Computing the source harmonic amplitudes becomes a cumulative procedure based on computing the best adjacent harmonic amplitude ratios, starting with the 1st and 2nd harmonic and working up to the highest harmonic. Applying this procedure, based on analysis, for up to 19 harmonics for the data of Fig. 12b gives the new, optimally aligned overlapped result shown in Fig. 14.

The amount dB_k that each harmonic track is shifted upward vertically, when negated, is the implied source harmonic's decibel amplitude, i.e., $-dB_k$. This yields the estimated source spectrum graph shown in Fig. 15. This is compared to the ideal decibel source spectrum, 20 log(1/*k*).

As was done in Fig. 4, the violin filter function can be estimated by first sorting the Fig. 14 data in frequency order and then smoothing it with a rectangular window filter. The result shown in Fig. 16 is characterized by a number of resonances, notably at f = 280, 1040, 1690, 1970, 2050, 2520, 2950, 3720, 4220, 4710, and 6100 Hz.

5 Pérez et al. Violin Input-Output Measurement

In 2011 Pérez-Carrillo et al. [19] reported on their experiment to measure the transfer response of a violin. A bowed glide from G3 to G4 was executed over a 40 s period on a rigidly suspended violin by a human performer in an anechoic chamber (at Aalto University in Finland). The input signal was recorded from a Yamaha VNP1 piezoelectric pickup mounted on the violin's bridge, and output signals were recorded from a semi-spherical array of 21 microphones. The Pérez-Carrillo et al. group performed short-time Fourier analyses of the input and output signals and estimated the violin's body frequency response (BFR) by taking ratios of the corresponding output and input components. Recently Dr.

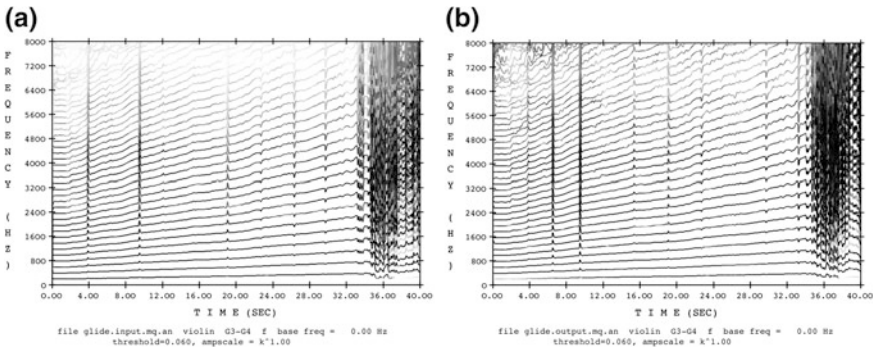


Fig. 17 Two-dimensional spectrograms of a G3-G4 violin glide tone. **a** Input bridge signal; **b** output microphone signal. *Vertical marks* are interruptions caused by bow reversals

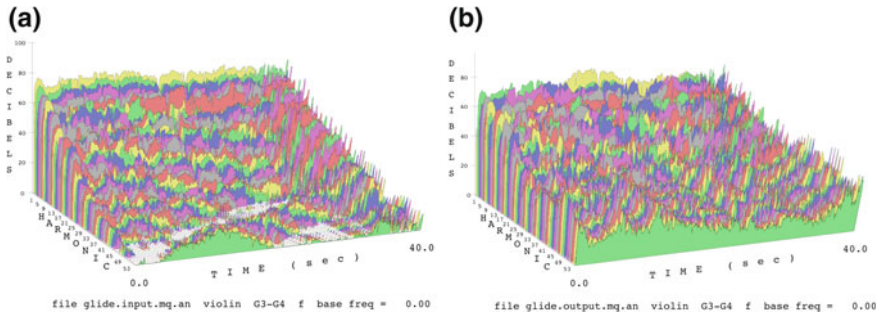


Fig. 18 Three-dimensional spectrogram of the G3-G4 violin glide tone showing amplitude-versus-time for harmonics 1–56. **a** Input signal recorded from bridge pickup; **b** radiated output recorded in an anechoic chamber

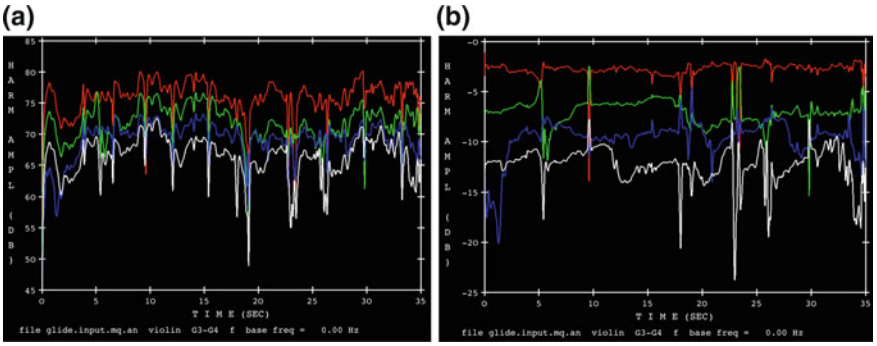


Fig. 19 For the violin G3-G4 glide tone input the amplitudes (in decibels) of harmonics 1 (*red*), 2 (*green*), 3 (*blue*), and 4 (*white*) are plotted versus time: **a** original dB amplitudes; **b** dB amplitudes with dB rms amplitude subtracted

Pérez-Carrillo graciously provided me with a sample input and a corresponding output signal from one of their recording sessions.

This author analyzed these input and output signals using SNDAN's MQ frequency-tracking method, and the tracks were converted to harmonics using SNDAN's two-way-mismatch F0 detector and harmonic separator [22]. The input and output signal's harmonic frequencies for this glide tone are shown in the 2-dimensional spectrogram graph in Fig. 17.

One question is: How variable is the source spectrum over time? As shown in Fig. 18a, at least for this recording, it is quite variable. As expected, the output spectrum, as shown in Fig. 18b, is more variable because it reflects the resonances that are a result of the violin body's filtering action.

The temporal variability of the input spectrum picked up on the bridge is illustrated for the first four harmonics in descending order in Fig. 19, with Fig. 19a

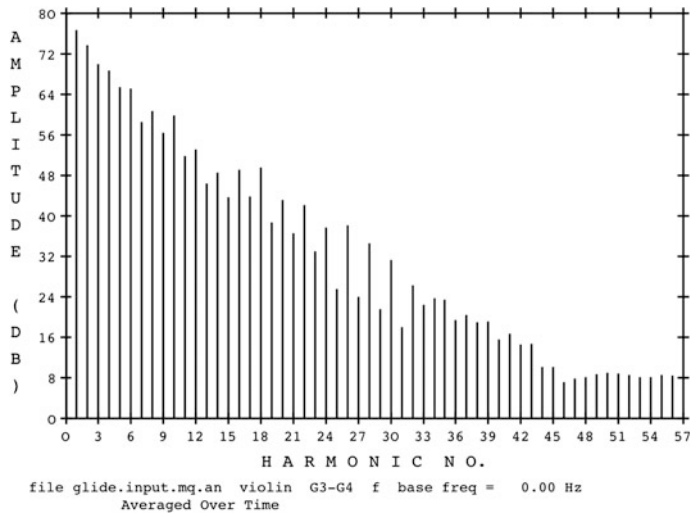


Fig. 20 G3-G4 violin glide tone average input signal harmonic amplitudes (in dB) versus harmonic

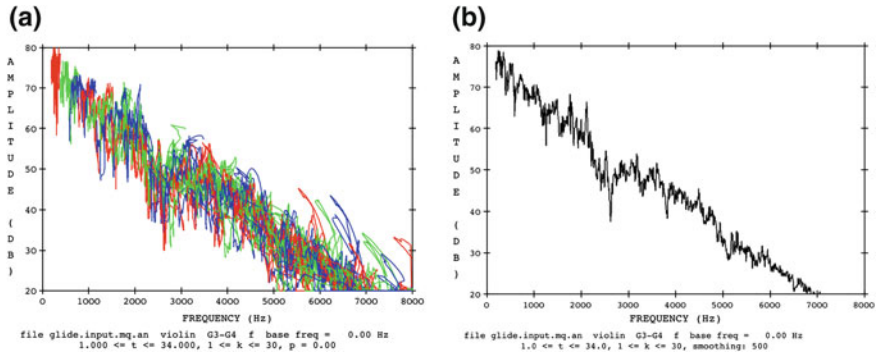


Fig. 21 G3-G4 violin bridge glide tone input signal spectrum: **a** MQ harmonic track analysis; **b** smoothed version of (**a**)

showing $dB_k(t)$ harmonic amplitudes and 19b showing $\{dB_k(t) - 20\log_{10}(A_{rms}(t))\}$. The first harmonic (in red) is almost flat in Fig. 19b because it closely follows the rms calculation. Note that in Fig. 19a, the curves approximately rise and fall together, indicating $dB_{rms}(t)$ having a significant effect on the harmonic amplitudes' behavior.

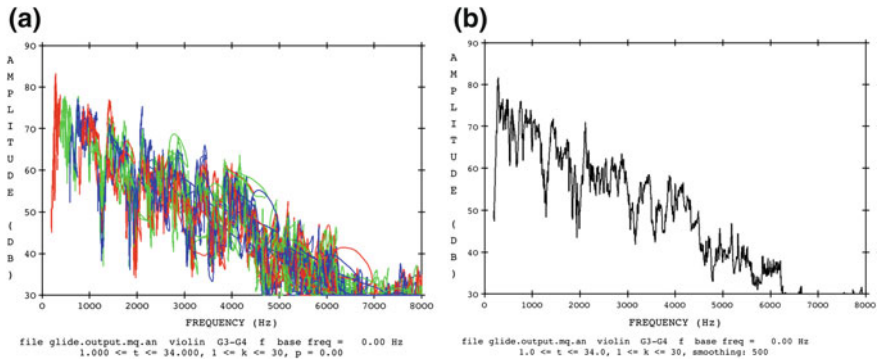


Fig. 22 G3-G4 violin radiated output glide tone output signal spectrum: **a** MQ harmonic track analysis; **b** smoothed version of (**a**)

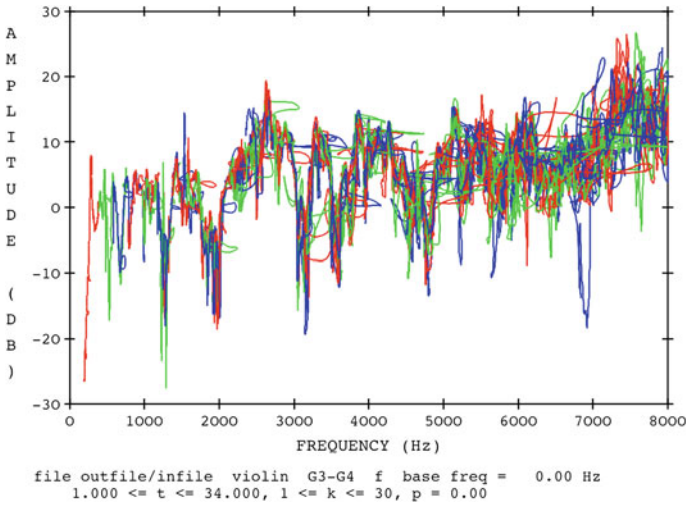


Fig. 23 G3-G4 violin glide tone transfer function computed by dividing output by input MQ harmonic analysis data. Thirty overlapped harmonics are used in this calculation

The downward trend of the input harmonic amplitudes continues as a function of harmonic number. Averaging them over time (using linear amplitude averaging) gives the spectrum (in decibels vs. harmonic number) shown in Fig. 20. There actually seem to be two superimposed downward trends, one which is about −1.9 dB per harmonic and the other which is about −1.5 dB per harmonic. Neither of these follows the ideal 20log10(1/k) spectrum. Rather it indicates an approximate exponential decay of the linear amplitudes.

The MQ frequency-tracking STFT method, which was used for the C5-C4 glide tone discussed in the previous section, was applied to both the input and output signals of the G3-G4 glide tone. The results are shown in Figs. 21 and 22. Note that

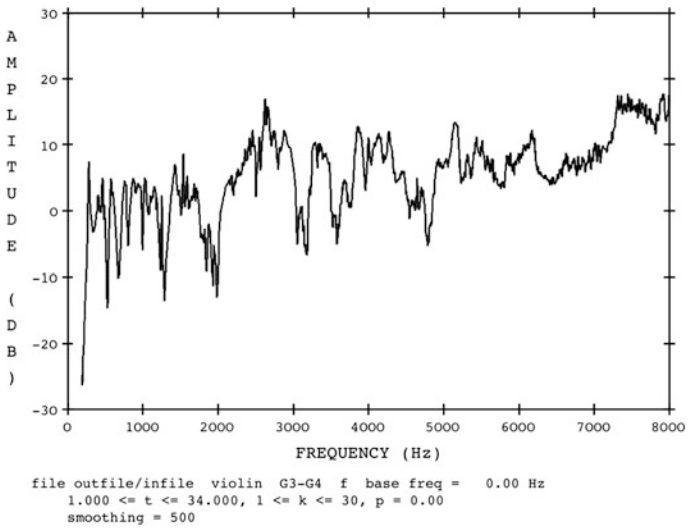


Fig. 24 G3-G4 violin glide tone transfer function computed by smoothing data of Fig. 23

the G3-G4 tone's track structures are much more complex than those of the C5-C4 tone (see Fig. 12) because the duration of the former is more than 10 times longer than the latter, so more time is spent analyzing each frequency region.

By taking ratios of the harmonic amplitudes for the input and output MQ analysis files for the G3-G4 glide tones, assuming they are synchronized in terms of frequency and time, we can compute the transfer function for the Pérez et al. violin. This is shown in Fig. 23, and the smoothed version is given in Fig. 24. These figures may be compared to Figs. 14 and 16, which are the corresponding graphs for the C5-C4 glide tone. Probably due to the much longer duration of the G3-G4 glide signals, the transfer function calculated from them shows much more detailed resonant behavior—analysis channels can be much narrower in frequency. The main difference, however, is the upward trend of the G3-G4 filter curve which shows no tendency to roll off before 8000 Hz as the estimated result for the C5-C4 tone does.

6 Discussion and Conclusions

The fact that overlaid harmonic amplitude-versus-frequency plots for operatic-style singing voices (at least for male voices) describe resonance characteristics that can lead to a general theory for simulating voice tones using a source/filter (aka subtractive synthesis) model leads one to be optimistic about the possibility of extending this idea to include various sustained-tone instruments. Because the violin, like the

voice, consists of two distinct physical parts, the string and body (analogous to the voice's glottis and the vocal tract) that are very independent, the violin would seem to be a good candidate for this process. However, violin vibrato depths (frequency deviations) tend to be much smaller (approx. ±0.6 %) than those typical of the operatic voice (approx. ±6 %). Therefore, the vibrato patterns of adjacent violin harmonics do not tend to overlap until very high harmonics are reached. Even so, the violin's small vibrato depths typically cause large harmonic amplitude modulations (see Fig. 10), indicating the presence of narrow resonance structures.

Therefore, to investigate a source/filter system for the violin, it seems better to use a bowed glide tone. Such a tone is easy for a human player to perform on the violin because, unlike the case for wind instruments, there are no constraints that make it difficult to "play between notes".

This paper explored two different attempts to use glide tones to separate source from filter. The first method discussed herein, which used the MQ time-variant spectral analysis algorithm, was based on the assumption that there exists a source which has a fixed waveform (and thus a fixed spectrum) and amplitude over the extent of the glide. This is obviously not true because amplitude-versus-frequency tracks of the various harmonics do not line up perfectly. On average the standard deviation between adjacent harmonics was 4.4 dB, and the average correlation between these tracks was 0.63.

With the second method two separate violin glide signals were analyzed again with the MQ method, one which was taken from a violin bridge and could be considered to be the source signal, and the other which was taken from the corresponding radiated sound in an anechoic chamber [19]. Here it is seen that the bridge harmonic amplitudes do vary significantly (see Figs. 18 and 19), but this could be because of resonances in the bridge admittance or non-uniform bowing pressure in the performance. Another result of note is that the amplitudes of the harmonics from the bridge, averaged over time, follow a spectrum that decays faster, with respect to harmonic number, (see Fig. 20) than the ideal $20\log_{10}(1/k)$ spectrum. Over a span of 18 harmonics, while the first method gives a 31 dB decrease in amplitude in the source spectrum, the second method averaged between low and high trends yields a 29 dB drop over the same number of harmonics.

However, the transfer functions for the two methods look quite different. For the same amount of smoothing, the second (G3-G4) method shows much more detail in the resonances than the first (C5-C4) method, undoubtably because the duration of the second's glide is much longer (34 s vs. 2.8 s). Also, the second method's transfer function shows a general trend upward of about 20 dB over a frequency range of 0–8000 Hz (see Fig. 24), whereas the first method's transfer function is flat up to about 4500 Hz and then begins to sharply roll off (see Fig. 16). It was found that the average spectral centroid of the first method's output spectrum (see Fig. 12) is 781 (std 255) based on 20 harmonics, whereas this average for the second (see Fig. 22) is 1216 based on 30 harmonics (std 225). Perhaps these differences can be attributed to the fact that the two methods used different violins played by different performers in different locations, and that the second violin or its performer or microphone location

are such that the second violin's output is simply much brighter than the first. More research needs to be done to reconcile these differences.

Acknowledgments The author is indebted to Ashwin Kalyan Vijayakumar for providing the C5-C4 glide tone and to Alfonso Pérez-Carrillo for sending the author the two files for the G3-G4 glide tone, as well as to Robert C. Maher for implementing the McAulay-Quatieri spectral analysis algorithm on the Unix platform. The vocal tone analyzed was collected by J. Charles O'Neil.

References

1. Seashore, C.E.: Psychology of music. McGraw-Hill, New York (1938)
2. Sundberg, J.: The science of the singing voice, p. 174. Northern Illinois University Press, Dekalb (1987). Fig. 8.13
3. Bretos, J., Sundberg, J.: Measurements of vibrato parameters in long sustained crescendo notes as sung by Ten Sopranos. J. Voice **17**(3), 343–352 (2003)
4. Maher, R., Beauchamp, J.: An investigation of vocal vibrato for synthesis. Appl. Acoust. **30**, 219–245 (1992)
5. Arroabarren, I., Carlosena, A.: Effect of the glottal source and the vocal tract on the partials amplitude of vibrato in male voices. J. Acoust. Soc. Am. **119**(4), 2483–2497 (2006)
6. Fletcher, H., Sanders, L.C.: Quality of violin vibrato tones. J. Acoust. Soc. Am. **41**(6), 1534–1544 (1967)
7. Mellody, M., Wakefield, G.H.: The time-frequency characteristics of violin vibrato: modal distribution analysis and synthesis. J. Acoust. Soc. Am. **107**(1), 598–611 (2000)
8. Saunders, F.A.: The mechanical action of violins. J. Acoust. Soc. Am. **9**(2), 81–98 (1937)
9. Hutchins, C.M.: Instrumentation and methods for violin testing. J. Audio Eng. Soc. **21**(7), 563–570 (1973)
10. Meinel, H.: Regarding the sound quality of violins and a scientific basis for violin construction. J. Acoust. Soc. Am. **29**(1), 817–822 (1957)
11. Mathews, M.V., Kohut, J.: Electronic simulation of violin resonances. J. Acoust. Soc. Am. **53** (6), 1620–1626 (1973)
12. Kohut, J., Mathews, M.V.: Study of motion of a bowed violin string. J. Acoust. Soc. Am. **49** (2), 532–537 (1971)
13. Schelleng, J.C.: The bowed string and the player. J. Acoust. Soc. Am. **53**(1), 26–41 (1973)
14. Raman, C.V.: On the mechanical theory of the vibrations of bowed strings and of musical instruments of the violin family with experimental verification of the results—part I. Bull. Indian Assoc. Cultiv. Sci. **15**, 1–158 (1918)
15. Schumacher, R.T.: Self-sustained oscillations of the bowed string. Acustica **43**(2), 109–120 (1979)
16. McIntyre, M.E., Woodhouse, J.: On the fundamentals of bowed-string dynamics. Acustica **43** (2), 94–108 (1979)
17. Lawergren, B.: On the motion of bowed violin strings. Acustica **44**(3), 194–205 (1980)
18. Lawergren, B.: Harmonics of S motion on bowed strings. J. Acoust. Soc. Am. **73**(6), 2174–2179 (1983)
19. Pérez-Carrillo, A., Bonada, J., Pätynen, J., Välimäki, V.: Method for measuring violin sound radiation based on bowed glissandi and its application to sound synthesis. J. Acoust. Soc. Am. **130**(2), 1020–1027 (2011)
20. O'Neil, J.C.: Computer analysis and synthesis of a sung vowel. DMA thesis, University of Illinois, Urbana-Champaign, IL, USA (1984)
21. Beauchamp, J.W.: Unix workstation software for analysis, graphics, modification, and synthesis of musical sounds. Audio Eng. Soc. Preprint No. 3479, pp. 1–17 (1993)

22. Beauchamp, J.W.: Analysis and synthesis of musical instrument sounds. In: Beauchamp J. W. (eds) Analysis, Synthesis and Perception of Musical Sounds: Sound of Music, pp. 1–89. Springer, New York (2007)
23. McAulay, R.J., Quatieri, T.F.: Speech analysis/synthesis based on a sinusoidal representation. IEEE Trans. Acoust. Speech, Signal Processing ASSP **34**(4), 744–754 (1986)
24. Beauchamp, J.W., Vijayakumar, A.K.: Estimation of a violin source/filter model. J. Acoust. Soc. Am. **137**(4, Pt. 2), 2405 (2015)
25. Maher, R.C., Beauchamp, J.W.: Fundamental frequency estimation of musical signals using a two-way mismatch procedure. J. Acoust. Soc. Am. **95**(4), 2254–2263 (1994)

Author Biography

James W. Beauchamp is a professor emeritus of the School of Music and a research professor and professor emeritus of the Department of Electrical and Computer Engineering, both at the University of Illinois at Urbana-Champaign, Urbana, Illinois in the USA. He is a fellow of the Acoustical Society of America (ASA) and of the Audio Engineering Society (AES). In addition he is a past chair of the ASA's Technical Committee on Musical Acoustics (TMU) and a past president of the International Computer Music Association. Most recently he served as an associate editor of Proceedings of Meetings on Acoustics (POMA) for the ASA and as the TCMU's representative to the ASA Membership committee.

Vowel Quality in Violin Sounds—A Timbre Analysis of Italian Masterpieces

Robert Mores

Abstract This chapter proposes the use of vowel quality (VQ) as a perception-based timbre descriptor for violin sounds. When asked to imitate a pitched sound with their own voices, listeners are likely to match an appropriate vowel to what they hear. The applied vowel represents a rough image of what is perceived by listeners, and the related VQ can be used as an intelligible timbre descriptor. The corresponding automated extraction of VQ from sounds follows the state of the art in speech analysis; however, the methods are amended to process stringed sounds. The automated VQ extraction agrees with subjective assessments. The objective is to assess violins by using their best sound examples, which are usually achieved in professional live performances. The impact of different professional musicians on variations in VQ is found to be low. The same holds for the impact of different recording distances in a given room. The investigated Italian masterpieces reveal the quality of front vowels. VQ representations present similarities among violins from the same maker and differences among violins from different makers. This result can also be heard.

1 Introduction

1.1 Aim of the Study

"Voices and strings: Close cousins or not?" Askenfelt [1] asked. He considered their comparable source-filter models and similarities in performance characteristics, such as vibrato, modulation, and pitch control. The question is valuable because sound quality is often assessed by using terms such as /a/-like, /o/-like, nasal, or singing. Musicians in particular tend to follow the voice paradigm when making assessments [28]. Although the list of attributes of voices is long–pitch, vibrato, nasality, sex, age, and more–this study concentrates on one attribute only:

R. Mores (✉)
University of Applied Sciences Hamburg, Finkenau 35, 22081 Hamburg, Germany
e-mail: Robert.Mores@haw-hamburg.de

A. Schneider (ed.), *Studies in Musical Acoustics and Psychoacoustics*,
Current Research in Systematic Musicology 4,
DOI 10.1007/978-3-319-47292-8_8

vowel quality (VQ). The VQ describes a vowel that is produced by a human vocal tract using two primary key parameters: tongue backness and tongue height, which accompanies jaw opening. The VQ should not be misunderstood as a measure of the quality of a particular vowel. The tongue position in continuous space initially translates to the VQ in continuous space. Subsequently, the perceived VQ might be limited to a few distinct vowels within a discrete space owing to categorical listening in a given language context.

The investigation bases on the observation that humans can usually imitate each others voice and very often also sounds from everyday life. When asked to imitate pitched sounds from musical instruments, listeners are likely to adjust the pitch and add a suitable vowel and some nasality, if appropriate. This has been observed with children and adults and encourages this investigation. Let us assume that listeners locate their tongue intuitively for the best fit between the produced vowel and the perceived sound. Then the tongue position, represented by the VQ, holds much of what a listener perceives. This feature is intelligible, as it can be understood by everyone with normal hearing and speaking abilities.

This investigation proposes automatic VQ extraction from recorded violin sounds in order to gain a broader view of the VQ of violin sounds. For this purpose existing state-of-the-art speech processing methods are applied and modified to process stringed sound. It is not claimed that violins naturally produce VQs nor that VQ would fully capture violin timbre. However, because VQ captures timbre to at least some extent, it makes sense to assess the Italian masterpieces from this point of view. This perspective, even if limited, has value because such a timbre description is directly linked to human perception and therefore to an intelligible representation.

1.2 Basics of Voices and Violins

As outlined by Askenfelt [1], the voice and the violin have in common that they produce periodic signals which will pass a system with few to many resonances. The vocal tract is generally considered as a duct with a cross section varied by the tongue, which determines the few dominant formants at low frequencies (see the analysis in Sect. 2.3 for examples of formant structures). The tongue position strongly determines resulting formants and related VQs.

Formants F1 to F3, in combination with the fundamental frequency F0, embody the VQ [31]. Today, F1 and F2 are generally believed to be sufficient for VQ determination [5, 32]. For an overview of the frequencies of F1–F5 across various VQs, see Sundberg's study [41]. This work from Sundberg is particularly relevant here, as it distinguishes between speaking and singing: professional singers merge formants F4 and F5 to establish a "singing formant" of strong projection, independent of the VQ [41].

In contrast to the voice, violins reveal a multitude of body and air resonances. Figure 1 illustrates an example using the "Schreiber" Stradivari violin from 1712.

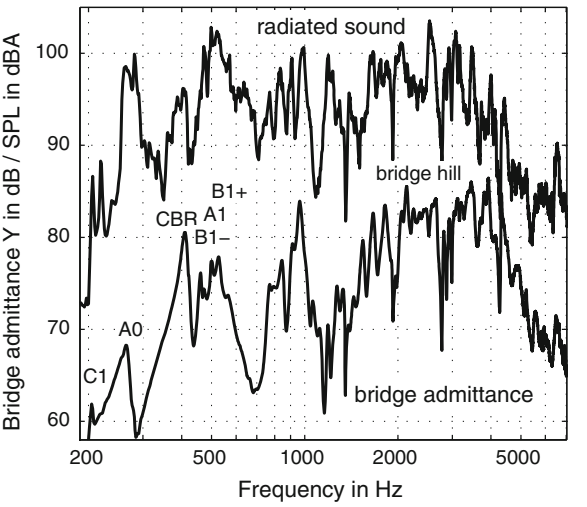


Fig. 1 Magnitude frequency response obtained by an impulse response analysis applied to the bridge of the "Schreiber" Stradivari, 1712. *Top* sound pressure level (SPL) in dBA averaged across 36 measurements around the violin, spaced by 10°, *bottom* bridge admittance *Y* in dB (arbitrary reference level)

The curve at the bottom is the magnitude frequency response obtained by a structural impulse response analysis applied to the bridge, called the bridge admittance. The impulse is a mini-force hammer impact at the G-string bridge side, in-plane. The impulse response is measured for the same direction, at the top of the bridge. The curve at the top of the figure represents the sound pressure level, or the radiated impulse response to the same hammer impact, averaged across 36 measurements around the violin within a semi-reverberant room. Both the radiated sound and the bridge admittance reveal a multitude of resonances. There are enough resonances in the low-frequency range where voice formants F1–F3 are usually hosted: (i) A0–A2, the Helmholtz and air cavity modes, (ii) the so-called signature modes, e.g., CBR, a body mode, and the B1− and B1+ modes, resulting from corpus bending and breathing, and (iii) other examples of body and plate modes up through 2000 Hz. Other resonances at roughly 3000 Hz compose the bridge hill.

Although both the voice and the violin can be studied to a first approximation by a source-filter model, the two are different. For the voice, the filter changes from vowel to vowel as the geometry and the resonances of the vocal tract are modified. These resonances are far apart and well separated, and the formants of the voice are univocally related to them. This is not the case for a violin; its body and therefore its resonances remain constant from note to note. Resonances are numerous and not far apart, and not all of them are excited by a given note. The pattern of a violin sound varies from note to note as different resonances are sampled by the note and its harmonic components. Thus, the difference between the formants of two vowels is due to a change in the geometry of the filter, whereas the difference between the formants of two violin sounds is due to a change in the source signal, and thus a modification of the pattern of excited resonances.

This explains why the formant structure of a radiated sound is not specific to a given violin alone but also specific to the individual notes that are played on that given violin. In combination with vibrato, formants will even vary within individual notes.

1.3 Related Works on Violin Sound Quality

Violin research is dominated by the measurement of physical features [18]. There is less work on violin sound quality, and the few existing studies concentrate on roughness or loudness [17], or on spectral bands' "fullness of sound" versus "brilliance" [7]. Meyer [26] studied the timbre of violins while relating the lowest air mode and the lowest plate modes to each other in terms of frequency and level. He showed that classification by timbre descriptors is possible, and he distinguished between violins made by Stradivari and Guarneri del Gesù as well as between Italian and non-Italian violins. His results are encouraging because this study seeks to make such classifications, albeit on the basis of timbre descriptors that are intelligible to musicians and luthiers. Fritz et al. [12] investigated tone quality on the basis of verbal descriptions and related these to acoustical properties or spectral bands, trusting that an analysis between physics and perception rather than examining physical properties alone delivers valuable insights.

More relevant to this study are timbre investigations that use analogies with the voice. Dünnwald explained the intensity and beauty of bel canto style in terms of a strong formant in the region of 3 kHz [14]. As mentioned above, Askenfelt generally discussed similarities between the human voice and strings and in particular compared their tone modulation potential [1]. Senn et al. [38] suggested a vowel analogy while quoting Meinel's definition of violin spectra within nine frequency bands [25]. This analogy comprises a sequence of phonetic sounds associated with the spectral shape of violins, from low to high frequencies: /u/, /o/, /a/, /ns/, /e/, /i/, and /s/. Tai and Chung [42] investigated 14 valued Italian and contemporary violins using the PRAAT speech analysis tool [4]. They identified front vowels in violin sounds and suggested a classification of male and female violin sounds.

Section 2 addresses the problems when applying state of the art speech processing tools to stringed sounds. The necessary adaptations are outlines in a section on signal processing. Section 3 shows some examples of how a typical multi-resonant violin sound can translate to VQ. Section 4 validates the core of the speech analysis tool on a technical basis, while the modifications towards sounds from bowed string instruments are validated by a perceptual test. Furthermore, separate studies on the impact of the musician and the impact of room acoustics are included here to understand how other factors might influence the extracted VQ while using sounds from performance. Section 5 contains the results for Italian masterpieces, and Sect. 6 concludes the investigation.

2 VQ Analysis Tool Preparation

2.1 *Applying Speech Analysis Methods to Bowed String Instruments*

Several issues arise when speech analysis methods are applied to sounds from bowed string instruments. First, speech analysis is often based on linear predictive coding (LPC); see Makhoul [23] for an introductory tutorial. The usually suggested order of 13 [11, 33] in LPC analyses is sufficient to describe up to five formants contained in the voice. Stringed sounds embody several dozens of resonances, and it seems inadequate to analyze them with an order-13 LPC. However, the goal is not to investigate the violin resonances as such or to achieve spectral resolution. The target is an automated extraction that would deliver formants and finally determine the VQ that would match assessments obtained from listening sessions. The LPC potentially solves this because the incorporated least-squares approximation seeks the best fit to a signal with a given spectral power distribution [29]. The analogy to the listening session is that listeners, as they seek a vowel within a sound, will be guided by the same spectral power distribution to identify one or the other VQ. Voice signals contain only a few well-separated formants with little room for ambiguity, whereas violin signals contain many spectral peaks and not necessarily an obvious formant. The relatively low order of the LPC effectively averages across spectral peaks of the stringed sound, and several peaks will finally constitute a formant in places where the density and level of peaks is relatively high. In summary, it is not recommended to optimize the parameters of the LPC analysis to understand the formant population of violins. LPC parameters are intentionally kept under the voice paradigm for the purpose of automatically extracting what a listener is likely to perceive in terms of VQ.

Second, the pitch of the male and female voice ranges roughly from 100 to 300 Hz, whereas that of the violin ranges roughly from 200 to 660 Hz for open strings and up to 2500 Hz in the treble range. The problem is the sparseness of harmonics which raises with an increasing fundamental F0. The spacing between two successive harmonics eventually becomes wider than the spacing between formants. Homomorphic deconvolution is an established method of separating the source and the filter [37] (tutorial in [29], Chap. 13) and has been successfully applied to high-pitched voices [34]. It potentially works for higher pitches and even for F0 > F1 because the working principle for separating the source and the filter is limited by the bandwidth of a low-pass filter in the intermediate cepstrum representation, which can be adapted to the task. Using homomorphic deconvolution prior to LPC analyses relaxes the problem of the sparse population of harmonics. This is particularly true for violin sounds, where bow friction causes prominent noise content in the signal, which is much more notable than the air jet contribution in voice signals. Thus, the violin is excited not only by more or less sparse

harmonics but also by broadband noise. Bow friction generates sufficient noise at low frequencies, which further encourages analyses for F0 > F1, which is unusual for speech. The generally experienced limitations for analyses of voices pitched higher than F1 [19] might therefore not strictly translate to analyses of violin sounds, especially when combined with homomorphic deconvolution.

Third, there is a minor issue of adaption. The spectral envelope of speech signals declines by roughly 12 dB/octave [10, 24]. Speech processing therefore often employs pre-emphasis of high-frequency components. Others use an additional two orders in LPC analyses for the spectral decline, which results in the widely used total order of 13 for five formants [11]. Sounds from violins, in contrast, have considerable energy up to 3 kHz before declining; see Fig. 1. Therefore, pre-emphasis and the extra pair of orders in LPC analyses are omitted, following the suggestions of Markel and Gray [24]. LPC is therefore used at order 11 without pre-emphasis.

2.2 Signal Processing

Automated formant extraction is based on well-established speech processing methods. After downsampling to f_s = 11,025 Hz, 40 ms long Blackman-windowed sections of sound with a 20 ms overlap are analyzed. This window size is commonly used in voice analyses [33] and is used here as well to follow the voice paradigm. It is sufficiently short to facilitate analyses of fast sequences of music. There is no need for wider windows because the analysis does not target the spectral resolution of individual spectral peaks of the violin. A noise gate eliminates passages of silence. Homomorphic deconvolution is applied to separate the source, i.e., the pitch generation, from the filter ([29, 34, 37], Chap. 13). This step facilitates the analysis of high-pitched sounds, where the spread of the harmonics becomes wider than the spread of the formants. LPC analysis of order 11 now contains F1 and F2. LPC analyses, in general, deliver coefficients that can be considered as the coefficients of a recursive filter, with the underlying roots representing the formants in the complex frequency plane. The roots are identified by standard root-solving according to Atal and Hanauer [2]. Among the several entries in the complex plane, formants F1 and F2 must be identified but also separated because for some vowels, F1 and F2 tend to merge. F1 and F2 are separated by a bandwidth criterion related to the separation criterion identified by Kim et al. [21] who qualified F1/F2 extractions by a minimum spacing of 700 Hz between them when working on the TIMIT data corpus. In the course of many listening sessions, a minimum bandwidth of 1000 Hz for both F1 and F2 has been defined to achieve the best match between the automatically extracted VQ and the perceived VQ; see Sect. 4.2. In parallel, the fundamental F0, which is related to the pitch of the note, is extracted by autocorrelation [30]. Finally, F0 and formants F1 and F2 are converted to International Phonetics Association (IPA) chart representations. See Fig. 2 for the entire process.

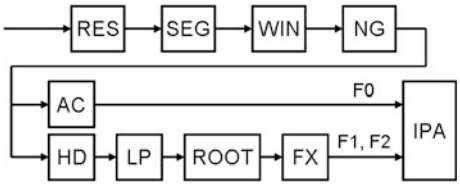


Fig. 2 Block diagram for automated formant extraction and tongue backness/height calculation. *RES* resampling to 11,025 samples per second, *SEG* segmentation to 442 samples = 40 ms, *WIN* Blackman windowing, *NG* noise gate, *AC* F0 extraction by autocorrelation, *HD* homomorphic deconvolution, *LPC* linear predictive coding of order 11, *ROOT* root solver, *FX* F1/F2 separation along bandwidth criteria, *IPA* conversion to IPA chart

2.3 *Representation of VQ on the IPA Chart*

The IPA chart is used to represent the VQ; see Fig. 3. The primary component is a measure of the tongue position, i.e. backness and height. The secondary component is the form of the lips, rounded or unrounded. For example, the vowel /e/ in "Easter" is located front close-mid with unrounded lips. This study considers only the primary component. Sound references are from the phonetic labs at UCLA. The IPA chart is preferred over an F1/F2 diagram for presenting the results, because it intuitively translates to the tongue position and jaw opening and is therefore more intelligible than F1/F2 plots. A reader can directly imitate the sound and imagine its stringed cousin. Another reason is that F0 also plays a minor role in VQ perception and a representation beyond F1/F2 capability is desirable [32, 43].

To translate from formant frequencies to the IPA chart, the height h and backness b are calculated by Pfitzinger's formula [32], which indicates the tongue position in an empirical way, without specifying physical dimensions:

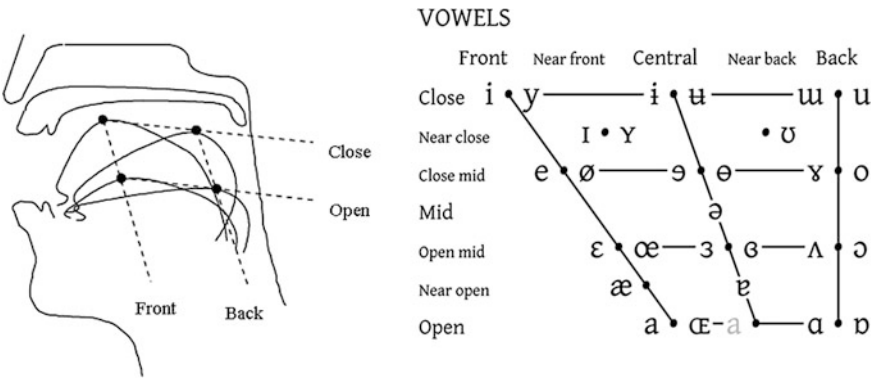


Fig. 3 Human vocal tract (*top*) and VQ in the IPA chart (*bottom*); vowels to right and left of dots are rounded and unrounded, respectively

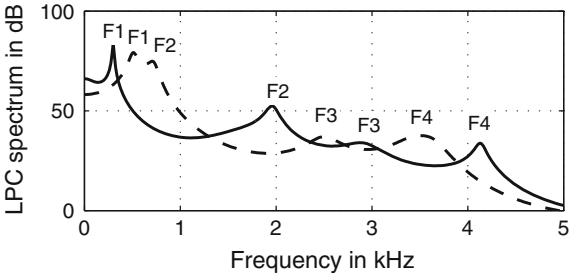


Fig. 4 Spectrum obtained from linear predictive coding (*LPC*) for male voice on a 700 ms section of sound, with typical formants indicated for two different vowels. *Solid* close front vowel, *dashed* open-mid back vowel

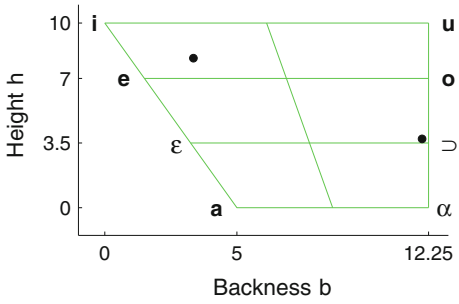


Fig. 5 IPA chart representation of an open-mid back vowel and a close front vowel spoken by male voice, both corresponding to the spectra in Fig. 4

$$h = 2.62 \cdot \log(F0) - 9.03 \cdot \log(F1) + 47.9$$
$$b = -0.49 \cdot \log(F0) + 1.74 \cdot \log(F1) - 8.39 \cdot \log(F2) + 59.2$$

Consider that the fundamental F0 forms part of the tongue height [16]. This can be observed in different ways: from /e/to /i/, the pitch naturally moves up, whereas it goes down when striving for an open-sounding /a/.

Figure 4 shows an example of LPC data for two different vowels, a close front vowel and an open-mid back vowel; the corresponding IPA chart representation is illustrated in Fig. 5. Note, that the formant frequencies rather than the formant levels are important for VQ perception.

3 Examples of VQ Extracted from Bowed String Instruments

A few examples might demonstrate how the spectral power distribution of an instrument translates to the perceivable VQ. Figures 6, 7 and 8 present the spectral analyses and extracted VQs for three different played notes. At the top of each figure, the spectrum of the coefficient sets obtained from LPC analyses is shown. Each trace represents the LPC spectrum of a 40 ms section of the entire sound sample. For comparison, the middle panel of each figure presents the power spectral

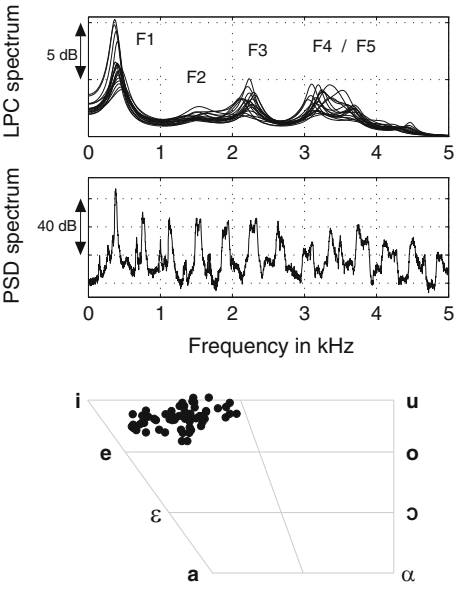


Fig. 6 G4 played on the "Willemotte" Stradivari for 1.8 s with some vibrato. *Top* spectrum of individual coefficient sets obtained from 40 ms Blackman-windowed sections of sound using linear predictive coding (*LPC*) of order $N = 11$ at a sample frequency of $f_s = 11{,}025$ Hz, *middle* power spectral density (*PSD*) for the entire sound, *bottom* IPA representation of formants, where each entry is strictly related to one of the individual coefficient sets from the 40 ms sections of sound

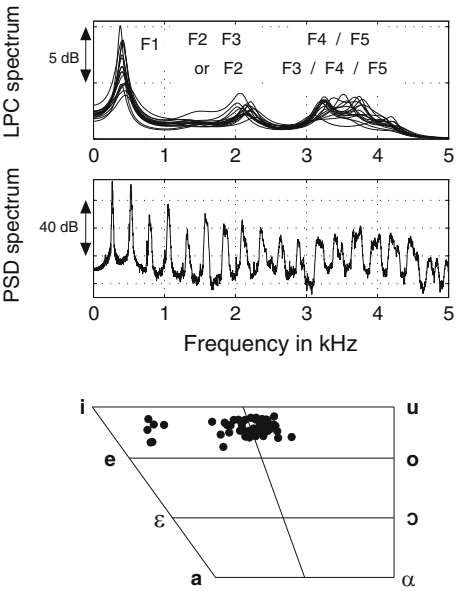


Fig. 7 C4 played on the "Willemotte" Stradivari for 1.7 s with some vibrato. *Top* spectrum of individual coefficient sets obtained from 40 ms Blackman-windowed sections of sound using linear predictive coding (*LPC*) of order $N = 11$ at a sample frequency of $f_s = 11{,}025$ Hz, *middle* power spectral density (*PSD*) for the entire sound, *bottom* IPA representation of formants, where each entry is strictly related to one of the individual coefficient sets from the 40 ms sections of sound

density (PSD) according to the well-known Welch method for the same sound but analyzed over the entire sound sample. At the bottom are the VQ representations related to the 40 ms sections. The sounds are individual notes taken from a scale that was played on the "Willemotte" Stradivari from 1734 [44].

For the data in Fig. 6, the tone G4 is played on the "Willemotte" Stradivari for 1.8 s with some vibrato. The analysis results are stable over the entire tone. A general observation for notes of such low pitch is that F1 is represented at a relatively strong level, in accordance with the existence of the signature modes that strongly radiate around 500 Hz; see Fig. 1. F2 and F3 are in the range of 1.5–2.5 kHz, where there are sufficient supporting air and plate modes in the violin. The perceptual match obtained when listening to this sound sample was convincing to the four members of the research team. This statement does not intend to proof anything at this point of the investigation, but it reflects the agreement of independent observations of expert listeners, including an internationally reputated luthier.

The reader may wish to compare this example with the voiced front vowel illustrated in Figs. 4 and 5. Positions in the IPA chart and related frequencies are similar. The F1 frequency is below 500 Hz and the F2 frequency is between 1.5 and 2 kHz for both sound sources. Back vowels, on the contrary, would appear with a collocation of F1 and F2 in the range below 1 kHz. The front vowel character is based on the signature modes around 500 Hz determining F1 and the plate modes in the range of 1 to 2 kHz determining F2, see Fig. 1.

Another likely outcome of the analyses is shown in Fig. 7, where another tone is analyzed from the same violin played by the same musician in the same recording session. C4 is played for 1.7 s, again with some vibrato. The general structure of the spectrum seems to be similar to that when the note G4 is played. However, the formant F2 shifted toward lower frequencies, effectively moving the VQ further toward back vowels. Additionally, there is only little energy in this frequency range, so there are a few traces without a peak between 500 Hz and 2 kHz. For a minority of sound sections, F2 will therefore be identified at approximately 2 kHz, where F3 is generally situated. This shifts the entry in the IPA chart toward front vowels. Determinations of either VQ can be caused by vibrato or directivity. The four listeners usually heard both vowels in such violin sounds, in a bi-stable fashion.

The analysis of individual tones across different violins typically reveals these two types of results, distinct or somewhat ambivalent VQ. In a systematic search among eight violins across different levels of quality and ranging from old to contemporary, 45 out of 120 sound samples reveal a distinct VQ [27]. It seems that high-quality violins tend to reveal a distinct VQ more often than low-quality instruments.

Other observations are:

(i) The strongly radiating signature modes around 500 Hz, together with the bulk of plate modes between 1000 and 2000 Hz, are likely to support front vowels rather than back vowels; see Fig. 4.

(ii) In the examples above, the formants F4 and F5 are typically located in the region between 3 and 3.5 kHz. Sundberg reports the same frequency range for F4 and F5 in singing voices and across a range of vowels [41]. Dünnwald

suggested that the range around 3 kHz is important for the bel canto character in the sound [7].

(iii) Formant F3 is identified across the range of 2–3.5 kHz in the above examples, depending on the results for F2, and in accordance with the findings of Sundberg. In the human voice, formant F3 is usually related to the structure of F1 and F2. In fact, F3 can be predicted from F1 to F2 [3]. The formant structure of F1 to F3 therefore contains inherent information on whether a sound suits the human voice analogy.

(iv) Amplitude differences between F1 and F2 play a role in vowel perception. In the context of a controversial discussion, most studies conclude that the level of F2 should be no more than roughly 30 dB lower than that of F1 before vowel perception will be misguided. The reader is referred to Kiefte et al. [20] for an overview and study. Here, the difference in levels is less than 30 dB; see the PSD spectra in Figs. 6 and 7.

4 Validation

4.1 Validation Against Voice Reference

First, the automated VQ extraction is referenced against voice signals. The implemented tool in its voice mode uses homomorphic deconvolution, LPC of order 13, and a first-order preemphasis, $H(z) = 1 - 0.94\ z^{-1}$. The performance of this analysis is validated against the Michigan vowel data corpus [15] and against the TIMIT data corpus using the related ground truth data of the MSR-UCLA VTR-Formant Database [6]. The TIMIT data corpus is of particular value because it contains 9981 nasals, which are usually considered a challenge to VQ extraction. Nasals are to be expected in violin sounds. The precision of the extracted frequencies for formants F1 to F3 across all utterances is higher than the precision achieved by PRAAT [4] and marginally lower than the precision achieved by WaveSurfer [39]. For details on the signal processing and on validation, see Smit et al. [40].

4.2 Analysis Modifications Toward Violins and Perceptual Verification

For analyses of violins, the formant extraction works with order-11 LPC and without pre-emphasis. The question is whether the tool will finally determine adequate VQ for violin sounds. Two steps were followed to gain confidence in the approach. First, a library of sampled normal- and high-pitched singing voices was used as a scaffold during listening sessions and tool development. Second, the final

tool was used to perceptually compare synthesized voice signals against sampled violin sounds across a wider pitch range.

A library of the singing voice was recorded and used as a reasonable intermediate reference for validation. Singing voice is still a voice signal, but has already something from the singing character of a string instrument. The library of singing voices contains two parts. Normal-pitched sounds are sampled from three women and one man aged 30 to 45 years across the A major scale ranging from A3 (220 Hz) to A4 (440 Hz), and across the German vowels /a/, /e/, /i/, /o/, and /u/. The high- pitched voice of a 10-year-old girl was recorded across the F major scale ranging from F4 (349 Hz) to A5 (880 Hz), and across the German vowels /a/, /e/, /i/, /o/, and /u/plus /œ/and /æ/. The library was used during listening sessions to empirically define bandwidth criteria. Analyses were done with order-11 LPC and without pre-emphasis.

The test in the second step examined whether the VQ perceived from voices would match the VQ perceived from violin sounds. Recordings were taken from 10 music students as they played chromatic scales on their own violins with various bowing techniques. The choice of violins was arbitrary and ranged from old to new and from moderate to good quality. From this library, a set of 15 legato non-vibrated sounds was chosen to cover a range of fundamental frequencies F0. The associated VQs were determined by automatic extraction; see the top of Fig. 8 for a representation in the IPA chart. Voice counterparts were synthesized using Fant's glottis impulse generator [9] together with shimmer and jitter [22]. For each violin sound, a set of 18 voice samples of varying VQ was synthesized using the pitch of the violin reference; see the bottom of Fig. 8 for the choice of discrete positions on the IPA chart. Twenty-six subjects with normal hearing abilities and one luthier voluntarily participated in a listening test in which they had to choose a voice sample that would preferably match the violin sample in terms of the VQ. For

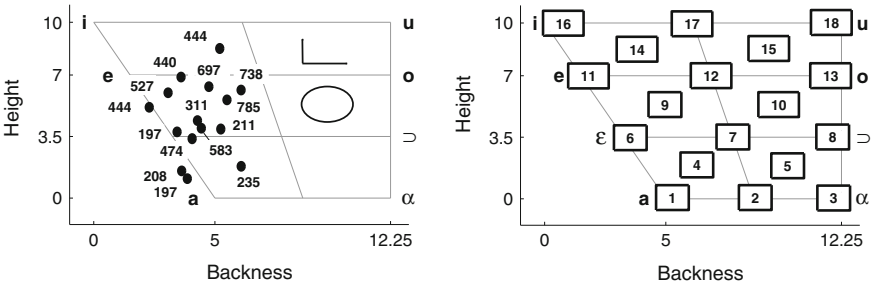


Fig. 8 IPA chart representations of voice/violin sound pairs for listening tests in which pairs had to be perceptually matched in terms of VQ. *Top* 15 samples of legato, non-vibrated sounds recorded from violins of moderate to good quality; *numbers* indicate the frequency of the fundamental F0 for each sound sample; *bars* indicate the evaluated mean distance between human response and automated VQ extraction for both tongue directions; *ellipse* indicates the standard deviation measured across all responses, *bottom* representation of voice sounds synthesized at the same fundamental frequency but with 18 different VQs

each task, the subjects were allowed to repeatedly play the violin sample or any of the 18 voice samples.

In the perceptual test, subjects predominantly chose a VQ near the automatically extracted VQ of the violin. The distance between the average response and the automatically extracted VQ is 1.76 units for backness and 1.29 units for height, averaged across all 15 voice/violin pairs; see the bars at the top of Fig. 8. The distances are larger for pairs pitched over 600 Hz than for pairs at lower pitches. The standard deviation across all responses is 2.17 units for backness and 2.13 units for height; see the ellipse at the top of Fig. 8. Please note that the average grid spacing between synthetic voice samples, 2.44 for backness and 1.67 for height, already significantly contributes to inaccuracy.

Subjects rated the VQ of the violin sound samples in close proximity to the automatically extracted VQ. Therefore, the analysis tools can be used as a predictor for human perception, given the indicated inaccuracy, and given the limited scope of pitches. The validation is also limited to front vowels, since the arbitrarily chosen violins did not embody back vowels.

4.3 Impact of Musician

Violin VQ is preferably extracted from sound samples of professional performance for good reason. A good violin played by a skilled musician in an acoustically adequate room corresponds to the preferred listening environment. A professional musician will produce the best sound examples of what a violin can do in a given acoustical environment. However, extracting the VQ from performance raises the question of how variations in individual performance relate to differences between violins.

A separate study addresses this influence of the musician. Two violins (VA: "Schreiber" Stradivari, 1712 and VB: a Saxonian mid-level instrument) were played by two professional musicians, MA and MB, in a luthier's studio ($V = 60\ \mathrm{m}^2 \times 4\ \mathrm{m} = 240\ \mathrm{m}^3$, $T_{60} \approx 0.5$ s, mono recording at 1 m distance). The music samples are first 13 bars of the Presto of the Sonata, BWV 1001 of J. S. Bach, corresponding to 7–9 s of performance (PRESTO), and nine bars of the second movement of the Sonata for Violin and Piano in A Major by C. Franck, corresponding to 17–19 s of performance (SONATA). See Fig. 9 for the music. None of the sounds were recorded on purpose to assist this study; the musicians were not advised to perform in a specific way, and have never met each other or heard each other's performance. The sound samples were chosen randomly from existing recordings; the only requirement was that they be of acceptable recording quality and facilitate pair-wise comparisons of musicians and violins for different pieces of music.

Figure 10 shows the empirical cumulative distribution functions of the F2 frequencies on the basis of several hundred analyzed 40 ms sections of sound. Given the 20 ms of overlap, there are some four analyzed 40 ms sections per note in the

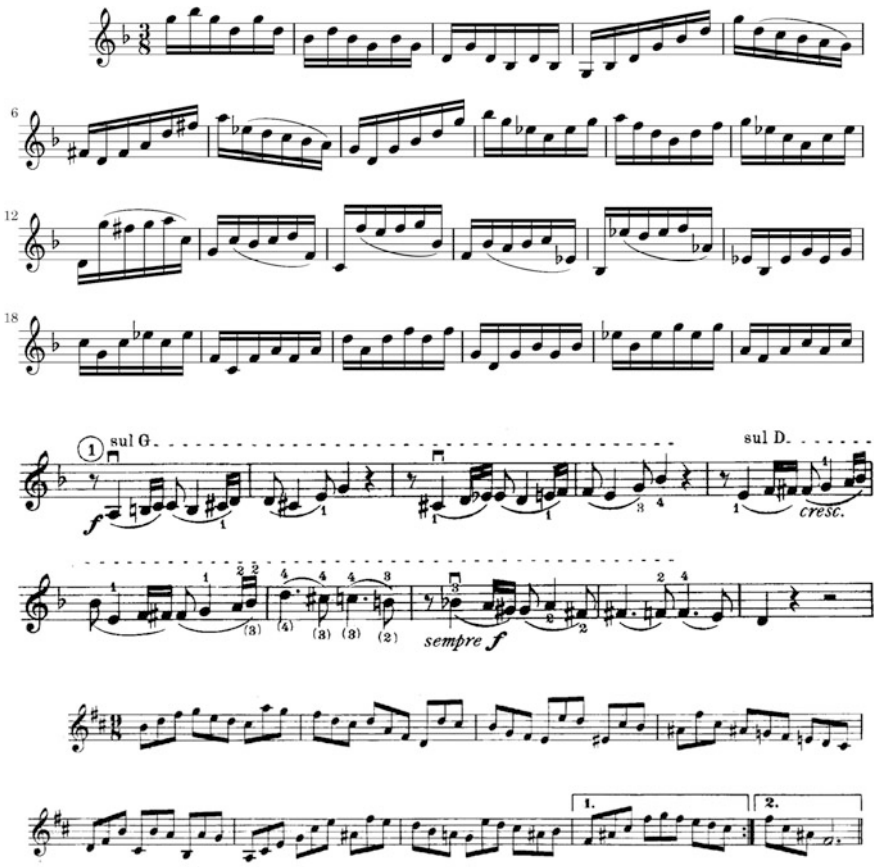


Fig. 9 Music played in violin performances used in qualifying studies. *Top* J. S. Bach, Sonata BWV 1001, Presto (PRESTO), *middle* C. Franck, Sonata for Violin and Piano in A Major, second movement (SONATA), *bottom* J. S. Bach, Double from Partita No. 1, BWV 1002 (DOUBLE)

fast PRESTO. Analysis sections that contain parts of two consecutive notes were sorted out by the pitch detector. In total, there are some 200 sections for the PRESTO and some 400 sections for the SONATA. Each trace in the figure represents one combination of violin, musician, and music played. The different pieces of music cause larger differences between the distributions than the musicians. However, the largest difference is caused by the difference between violins. The F1 distributions of the two violins are similar and are therefore less suited to investigate robustness issues. The robust F2 distribution is not a general proof and is only an example analysis, but it encourages feature extraction from performance. This example agrees well with the experience that there are usually noticeable timbre differences between violins, whereas professional musicians tend to perform in a

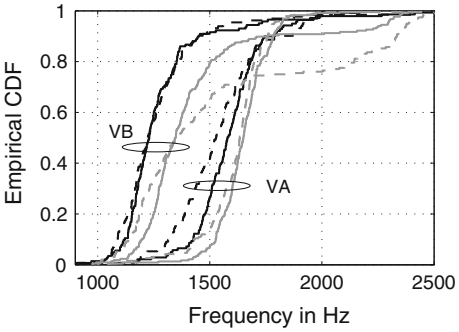


Fig. 10 Empirical cumulative distribution functions (*CDFs*) of F2 frequency. F2 is automatically extracted from some 200 40-ms sections of the PRESTO (*black*) and SONATA (*gray*) sound samples, played by two musicians (MA, *solid lines*; MB, *dashed lines*), on "Schreiber" Stradivari 1712 (VA) and a Saxonian mid-level violin (VB)

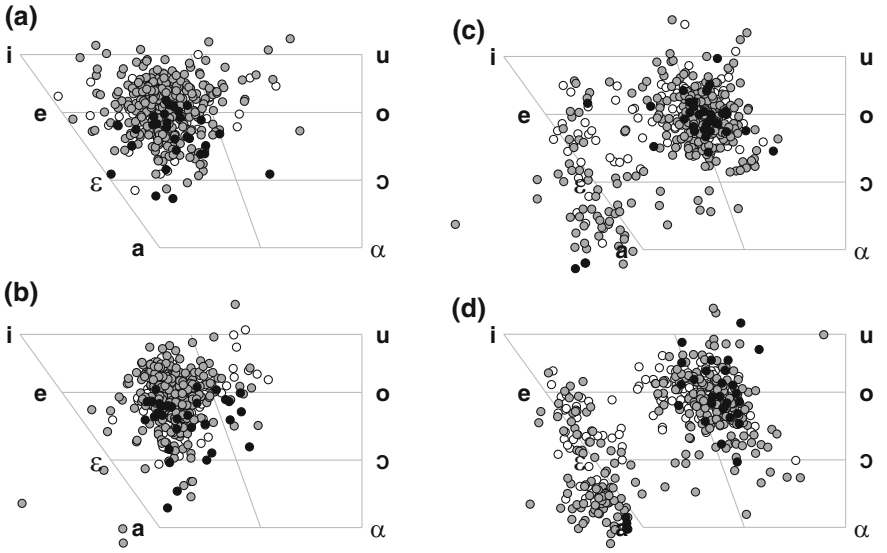


Fig. 11 IPA chart representations of F0, F1, and F2 automatically extracted from some 400 40-ms sections of SONATA sound samples. **a** Musician MA and **b** musician MB on the "Schreiber" Stradivari 1712 (VA), **c** musician MA and **d** musician MB on a Saxonian mid-level violin (VB). *Black circles* note played between G3 and C#4, *gray circles* note played between D4 and G#4, *white circles* note played A4 or above

very similar way. Admittedly, the VQ extracted from a performance will ultimately reflect the violin, musical piece, and musician. However, the objective here is to extract the VQ under common listening conditions rather than to precisely distinguish a musician and his instrument. This random example demonstrates that

although the distinction between violins might not be perfect, it can be obtained while exploiting the advantages of using sound examples from performances.

Figure 11 represents the VQ in the IPA chart derived from F0 to F2 for the two musicians on the two violins. The difference between violins is easy to recognize, whereas the difference between musical performances is less distinct.

4.4 Impact of Room Acoustics

Another set of data was investigated to understand the impact of the room and the recording. The question is whether varying player positions or movements would hamper comparative analyses of violins. This second data set consists of AB-stereo recordings originally used to understand the perceived distance versus the physical distance in semi-reverberant rooms. A professional musician played eight bars of the Partita No. 1 of the Double, BWV 1002 of J. S. Bach, corresponding to 20 s of performance (DOUBLE) on two violins (VC: Schleske Opus 96, 2008, VD: Markneukirchen student-level instrument, about 1900) in a university laboratory (V = 325 m^2 × 3 m = 975 m^3, $T_{60} \approx 0.75$ s). The musician played at different locations L1 to L4 (6.4, 8.2, 13, and 16 m away from the microphones, all locations outside the hall radius) and maintained the same style of playing. From the total library of 480 recordings 16 were chosen, 2 for each location and each violin. The only selection criterion was a preference for minimal environmental noise in the recording.

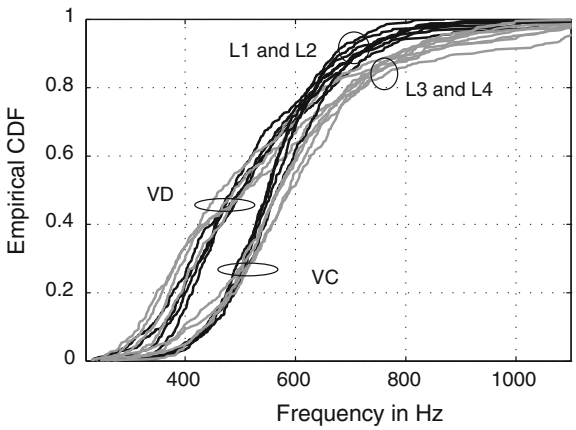


Fig. 12 Empirical cumulative distribution functions (*CDFs*) of F1. F1 was automatically extracted from roughly 1000 40-ms sections of sound samples of DOUBLE, played on Schleske Opus 96 (VC) and on a student-level instrument from the area of Markneukirchen (VD) at playing locations 6.4 m and 8.2 m (L1 and L2, *black*) and 13 m and 16 m (L3 and L4, *gray*) from the microphones. For each violin and each location, two arbitrary sound samples were chosen from a larger library

The two violins reveal differences for the extracted F1 distributions while F2 distributions are similar. Is the classification potential of F1 distributions likely to be hampered by varying playing positions? Figure 12 shows the empirical cumulative distribution functions of F1 based on roughly 1000 40-ms sections of sound, two for each violin and each location. The differences in the distribution functions are caused primarily by differences between the violins and secondarily by different playing locations. Within a given room, the formant structure is quite stable, although the timbre and therefore the associated VQ might deviate between rooms.

5 Results for Italian Masterpieces

5.1 *Investigated Recordings*

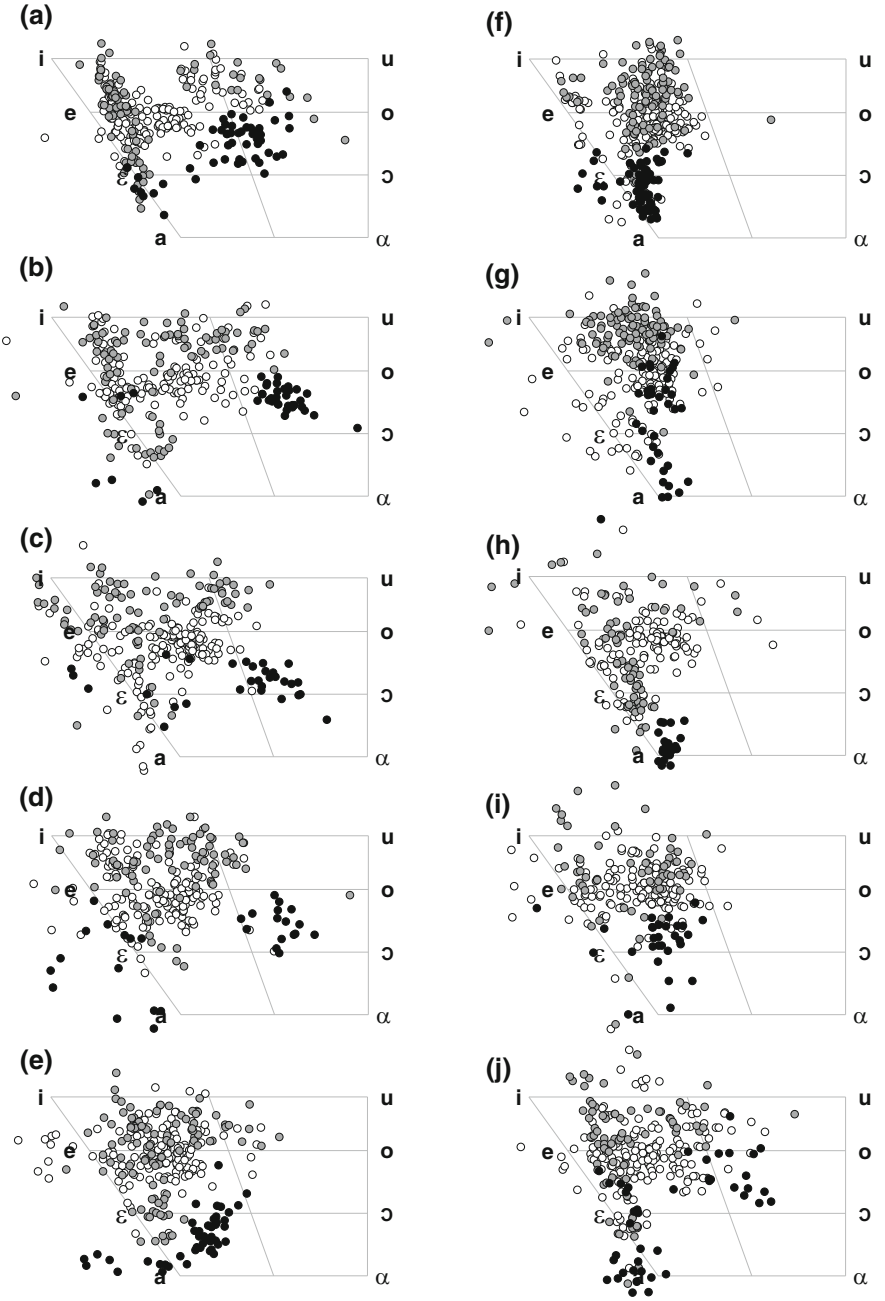
The following recordings were investigated:

(i) Ricci, 15 old Italian violins [35]
(ii) Ricci, 18 contemporary violins [36]
(iii) Strad3D, three valued old Italian violins [44]
(iv) Gawriloff, six old violins and one contemporary violin [13]
(v) Ehnes, nine old Italian violins [8].

In this paper only the results from (i) are presented. One of the reasons is that the recording is well engineered and balanced even though it is the oldest. The violins can be distinguished well, and we believe that the room has little impact on the timbre. The other recordings are less well engineered, and in some recordings the timbre of the room is too dominant.

5.2 *Results on Italian Masterpieces*

This section compares two Amati, five Stradivari, five Guarneri and another two valued Italian instruments. The music played is an excerpt from the violin concerto in G minor by Bruch [35]. The list of instruments is (order of tracks on the recording): a violin from Andrea Amati dated between 1560 and 1570, one from Nicolo Amati from 1656, “Spanish” Stradivari from 1677, “Ernst” Stradivari from 1709, “Joachim” Stradivari from 1714, “Monasterio” Stradivari from 1719, “Madrileno” Stradivari from 1720, “Rode” Stradivari from 1733, a violin from Gasparo da Saldo dated between 1570 and 1580, “Constable” Bergonzo 1731, “Gibson” Guarneri del Gesù from 1734, “Lafont” Guarneri del Gesù from 1735, “Plowden” Guarneri del Gesù from 1735, “Ex.-Vieuxtemps” Guarneri del Gesù from 1739, and “De Beriot” Guarneri del Gesù from 1744. Figure 13 compares the


(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
(j)
i
u
e
o
ɛ
ɔ
a
ɑ

◀ **Fig. 13** VQs of Guarneri del Gesù (*left column*) and Stradivari (*right column*) violins automatically extracted from 40 ms sections of sound from 24 s of Bruch's G-minor concerto. **a** "Gibson" 1734, **b** "De Beriot" 1744, **c** "Plowden" 1735, **d** "Lafont" 1735, **e** "Ex.-Vieuxtemps" 1739, **f** "Madrileno" 1720, **g** "Joachim" 1714, **h** "Monasterio" 1719, **i** "Spanish" 1677, **j** "Ernst" 1709. *Black circles* note played between G3 and C#4, *gray circles* note played between D4 and G#4, *white circles* note played A4 or above

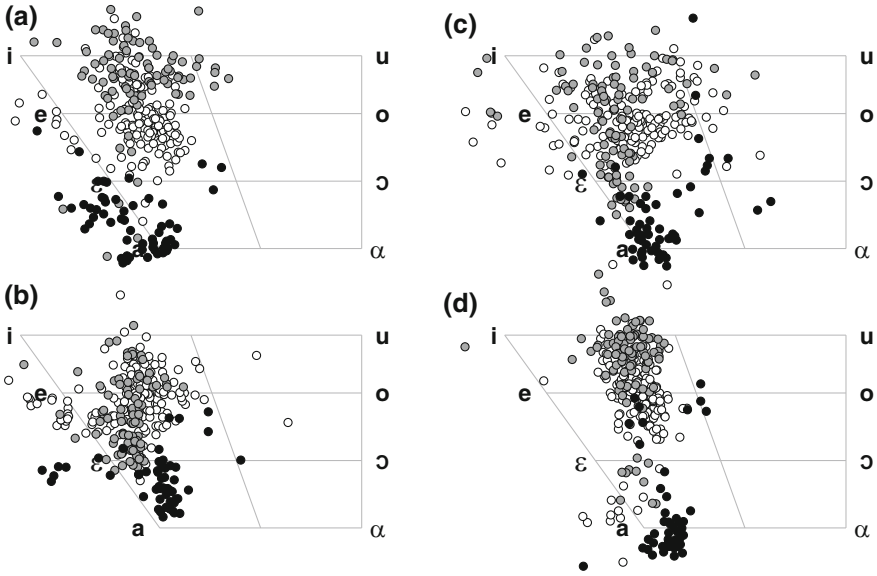


Fig. 14 VQs of valued Italian violins automatically extracted from 40 ms sections of sound from 24 s of Bruch's G-minor concerto. **a** Violin from Andrea Amati dated between 1560 and 1570, **b** violin from Nicolo Amati from 1656, **c** "Constable" Bergonzo 1731, **d** violin from Gasparo da Saldo dated between 1570 and 1580. *Black circles* note played between G3 and C#4, *gray circles* note played between D4 and G#4, *white circles* note played A4 or above

VQ of five Guarneri del Gesù violins against those of five of the Stradivari violins, and Fig. 14 shows the VQs of the four other Italian masterpieces.

The observations are:

(i) For all the violins, the bulk of the VQ representations reside well within the space of the IPA chart.

(ii) Violin sounds contain front vowels rather than back vowels. This finding is in accordance with the findings of Tai and Chung [42]. Recall Figs. 4 and 5 for typical differences between front and back vowels.

(iii) The differences between violins seem to be much stronger for lower notes than for higher notes. See the similarity of VQ populations for high-pitched notes (white circles) in Figs. 13e–j and 14.

(iv) There are similarities among violins of the same maker and differences between makers. For instance, the VQ representations of the top 4 (a–d)

Guarneri violins have many similarities; in particular, the "Gibson," the "De Beriot," the "Plowden," and the "Lafont" are similar. Likewise, the VQ representations of the five Stradivari violins are similar; in particular, the "Madrileno," the "Joachim," the "Monasterio," and the "Spanish" reveal the front vowel character more clearly than the Guarneri violins except the Vieuxtemps. The differences between the two makers are most obvious for the lowest notes played. Notes played on the G-string are represented by front vowels for Stradivari violins and by mid-to-back vowels for the Guarneri violins. Likewise, the notes played on the D-string are compactly represented by mid-to-front vowels for the Stradivari violins, whereas for the Guarneri violins the representation ranges from front to mid. The similarity between the "Ernst" Stradivarius and either the "Plowden" or "Lafont" Guarneri seems to be an exception. In summary, the violin makers are likely to be distinguishable by the VQ population.

(v) The VQ populations of the other four Italian violins in Fig. 14 are different from most of the Guarneri and Stradivari violins and different from each other, especially when considering the notes in the middle register (gray circles). Similarities are observable between Figs. 13f and 14b, and between 13j and 14c. The violin from da Saldo clearly differs from all the others. This result can also be heard.

6 Conclusions

Standard voice analysis methods have been extended to process sounds from bowed string instruments and to extract vowel quality (VC). The automated VQ extraction agrees well with subjective assessments of violin sounds for the range of front to mid vowels and for pitch frequencies up to 800 Hz. Analyzing violin sounds by voice-conformant methods of speech processing will yield formant structures that are quite comparable to those found for the human voice. The structural modes of violins support F1 and F2 at roughly 500 Hz and between 1000 and 2000 Hz, respectively. The translation of these formant structures into VQ will most likely yield front vowels. The results of VQ analysis of violin sound samples generally agrees with the range of VQs in the human voice. Representations of the VQ in the IPA chart allow a compact illustration of the results that directly translates them to perceived timbre because the reader can directly apply an identified tongue position and imitate the sound. While extracting VQ from performance the player has a minor influence on the VQ population in the IPA chart whereas the choice of music has a major influence. Among the six violin sound libraries investigated, the most well-engineered and balanced recording is analyzed, covering some 14 Italian masterpieces. On the basis of the VQ representations, commonalities can be observed for violins from the same maker. Differences between violins from different makers are identified, in particular for notes played on the lower two strings.

A reference piece of music or a reference scale should be used for future general classifications.

Acknowledgments The author thanks Martin Schleske, Munich, for providing sound samples, and the German Federal Ministry of Education and Research, BMBF, for funding the violin project, reference no. AiF 1767X07.

References

1. Askenfelt, A.: Voices and strings: close cousins or not? In: Sundberg J., Nord L., Carlson R., O'Dowd S.C. (eds.) Music, Language, Speech and Brain. Proceedings of an International Symposium at the Wennex-Gren Center, Stockholm, Stockholm 1990, pp. 243–256 (1991)
2. Atal, B.S., Hanauer, S.L.: Speech analysis and synthesis by linear prediction of the speech wave. J. Acoust. Soc. Am. **50**, 637–655 (1971)
3. Barlow, M., Clermont, F.: A parametric model of Australian English vowels in formant space. In: 8th Australian International Conference on Speech Science Technology, Australian Speech Science and Technology ASSTA, Australia, Canberra (2000)
4. Boersma, P.: Praat, a system for doing phonetics by computer. Glot Int. **5**, 341–345 (2001)
5. Delattre, P., Liberman, A.M., Cooper, F.S., Gerstman, L.J.: An experimental study of the acoustic determinants of vowel color; observations in one- and two-formant vowels synthesized from spectrographic patterns. Word **8**, No. 3 (1952)
6. Deng, L., Cui, X., Pruvenok, R., Huang, J., Momen, S., Chen, Y., Alwan, A.: A database of vocal tract resonance trajectories for research in speech processing. In: Proceedings ICASSP, pp. 60–63 (2006)
7. Dünnwald, H.: Ein Verfahren zur objektiven Bestimmung der Klangqualität von Violinen (English: A procedure for objective assessment of violin sound quality). Acoustica **58**, 162–169 (1985)
8. Ehnes, J.: Hommage (DVD and Compact Disc, Onyx Classics), tracks 22–30 (2008)
9. Fant, G., Liljencrants, J., Lin, Q.: A four-parameter model of glottal flow. Technical report, Dept. for Speech, Music and Hearing, KTH Schweden (1985)
10. Flanagan, J.L.: Some properties of the glottal sounds source. J. Speech Hear. Res. **1**, 99–116 (1958)
11. Flanagan, J.L., et al.: Speech Analysis Synthesis and Perception, p. 103. Springer, New York (2008)
12. Fritz, C., Blackwell, A.F., Cross, I., Woodhouse, J., Moore, B.C.J.: Exploring violin sound quality: Investigating English timbre descriptors and correlating resynthesized acoustical modifications with perceptual properties. J. Acoust. Soc. Am. **131**, 783 (2012)
13. Gawriloff, S.: What about this Mr. Paganini? (Compact Disc, Tacet, Stuttgart), tracks 1–7 (1996)
14. Heike, G., Dünnwald, H.: Neuere Klanguntersuchungen an Geigen und Ihre Beziehung zum Gesang (English: Recent sound analyses on violins and their relation to singing). In: Auhagen W., Gätjen B., Niemöller K.W. (eds.) Systemische Musikwissenschaft - Festschrift Jobst Peter Fricke zum 65. Geburtstag. Peter Lang Verlag, Frankfurt, pp 235–244
15. Hillenbrand, J., Getty, L.A., Clark, M.J., Heeler, K.W.: Acoustic characteristics of American english vowels. J. Acoust. Soc. Am. **97**, 3099–3111 (1995)
16. Hellström, Å., Aaltonen, O., Raimo, I., Vilkman, E.: The role of vowel quality in pitch comparison. J. Acoust. Soc. Am. **96**, 2133–2139 (1994)
17. Houtsma, A.J.M.: Letter to the editor: "On the relationship of Sound Power and Loudness of Violins". CASJ, ISS 4.8, p. 4 (2003)

18. Hutchins, C.M.: Research Papers in Violin Acoustics 1975–1993 (Acoustical Society of America), p. 15 (1997)
19. Joliveau, E., Smith, J., Wolfe, J.: Vocal tract resonances in singing: the soprano voice. J. Acoust. Soc. Am. **116**(Pt. 1), 2434–2439 (2004)
20. Kiefte, M., Enright, T., Marshall, L.: The role of formant amplitude in the perception of/i/and/u/. J. Acoust. Soc. Am. **127**, 2611–2621 (2010)
21. Kim, C., Seo, K., Sung, W.: A robust formant extraction algorithm combining spectral peak picking and root polishing. EURASIP J. Appl. Sign. Process. **2006**, 1–16 (2006) Article ID 67960
22. Klatt, D.H.: Software for a cascade/parallel formant synthesizer. J. Acoust. Soc. Am. **67**, 971–995 (1980)
23. Makhoul, J.: Linear prediction: a tutorial review. Proc. IEEE **63**(4), 561–580 (1975)
24. Markel, J.D., Gray, A.H.: Linear Prediction of Speech, p. 154. Springer, New York (1976)
25. Meinel, H.: Regarding the sound quality of violins and a scientific basis for violin construction. J. Acoust. Soc. Am. **29**, 817–822 (1957)
26. Meyer, J.: Zum Klangphänomen der altitalienischen Geigen (English: On the sound phenomenon of old Italian violins). Acustica **51**, 1–11 (1982)
27. Mores, R.: Vowel quality in violin sounds. In: Bader R., Neuhaus C., Morgenstern U. (eds.) Concepts, Experiments, and Fieldwork: Studies in Systematic Musicology and Ethnomusicology. Peter Lang Verlag, pp 113–136 (2010)
28. Nykänen, A., Johansson, Ö., Lundberg, J., Berg, J.: Modelling perceptual dimensions of saxophone sounds. Acta Acustica United With Acustica **95**, 539–549 (2009)
29. Oppenheim, A.V., Schafer, R.W.: Discrete-Time Signal Processing, 3rd edn, p. 916. Pearson, Upper Saddle River (2010)
30. Peeters, G.: Music pitch representation by periodicity measures based on combined temporal and spectral representations. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP-2006), vol. 5, pp. 53–56 (2006)
31. Peterson, G.E., Barney, H.I.: Control methods used in the study of vowels. J. Acoust. Soc. Am. **24**, 75–184 (1952)
32. Pfitzinger, H.: Towards functional modelling of relationship between the acoustics and perception of vowels. ZAS Pap. Linguist. **40**, 133–144 (2005)
33. Rabiner, L.R., Schafer, R.W.: Digital Processing of Speech Signals, p. 420. Prentice Hall, New Jersey (1978)
34. Rahman, M.S., Shimamura, T.: Linear prediction using refined autocorrelation function. EURASIP J. Audio Speech Music Process. **2**, 1–9 (2007)
35. Ricci, R.: The glory of Cremona (Compact Disc, compilation by MCA Records), original recording from 1963, MCA catalog number SXSE 7179, tracks 1–15 (1989)
36. Ricci, R.: The legacy of Cremona (Compact Disc, Dynamic Srl, Italy), tracks 17–30 (2001)
37. Schafer, R.W., Rabiner, L.R.: System for automatic formant analysis of voiced speech. J. Acoust. Soc. Amer. **47**, 637–648 (1970)
38. Senn, W., Winckel, F., Winternitz, E.: Violine. In: Musikinstrumente in Einzeldarstellungen, Band 1: Streichinstrumente (English: Musical instruments in dedicated portrays, vol. 1: strings) (dtv/Bärenreiter, Kassel), pp. 21–22 (1981)
39. Sjölander, K., Beskow, J.: Wavesurfer-an open source speech tool. In: Proceedings International Conference on Spoken Language Processing, pp. 464–467 (2000)
40. Smit, T., Tuerckheim, F., Mores, R.: Fast and robust formant detection from LP data. ELSEVIER Speech Communication **54**(7), 893–902 (2012)
41. Sundberg, J.: Articulatory interpretation of the singing formant. J. Acoust. Soc. Am. **55**, 838–844 (1974)
42. Tai, H.-C., Chung, D.-T.: Stradivari violins exhibit formant frequencies resembling vowels produced by females. Savart J. **2** (2000)
43. Vaissiere, J.: On the acoustic and perceptual characterization of reference vowels in a cross-language perspective. In: 17th International Congress of Phonetic Sciences (ICPhS XVII), Aug 2011, China. pp. 52–59 (2000)

44. Zygmuntowicz, S., Bissinger, G.: Strad 3D–Scientists, violinmakers & musicians study 3 classic violins: Plowden, Willemotte, Titian (VSA Oberlin Acoustics), sounds extracted from disc 1 (2000)

Author Biography

Robert Mores works as a professor at the University of Applied Sciences in Hamburg where he teaches telecommunications, digital signal processing, and acoustics. He received a diploma in electrical engineering in 1988 and a Ph.D. in computer science from De Montfort University Leicester, UK, in 1994. His research is partially devoted to musical acoustics. As a musician, he plays the violin and the cello.

Sound, Pitches and Tuning of a Historic Carillon

Albrecht Schneider and Marc Leman

Abstract The City of Bruges in Flanders owns one of the finest carillons in Europe. Of its 47 bells, 26 are historic specimens, cast by Joris Dumery (Georgius Du Mery) between 1742 and 1748. In 2010/11, the carillon underwent restoration including retuning as necessary. The present article reports the status of the 26 historic carillon bells as recorded by us in the years 1997–2000 prior to restoration. Since the original tuning of the bells has been assumed to be close to quarter-comma meantone temperament, the tuning is investigated both in regard to physical data and scaling (weight, diameter) as well as fundamental frequencies and spectral characteristics of the Dumery bells. Trajectories for the five so-called principal partials hum, prime, tierce, fifth and octave (or nominal) are established to check the smoothness of inner tuning of the 26 bells. From the fundamental frequencies, the tuning of the 26 Dumery bells to a musical scale is derived, and a matrix of fundamental frequencies shows all intervals that can be realized with these bells. A second parameter relevant for the tuning of (swinging and carillon) bells is the so-called strike note, which is first discussed with respect to concepts of pitch perception and then in regard to a possible meantone tuning. Finally, in continuation of previous experiments which demonstrated ambiguity of pitch perception in subjects listening to bell sounds we conducted two small experiments one of which addresses the number of pitches subjects distinguish per bell sound while the other

A. Schneider (✉)
Institute of Systematic Musicology, University of Hamburg, Hamburg, Germany
e-mail: aschneid@uni-hamburg.de

A. Schneider
Institut für Systematische Musikwissenschaft, Neue Rabenstr. 13,
20354 Hamburg, Germany

M. Leman
Institute of Psychoacoustics and Electronic Music, University of Ghent, Ghent, Belgium
e-mail: Marc.Leman@ugent.be
URL: http://www.ipem.ugent.be

M. Leman
IPEM—Musicologie, Vakgroep Kunst, Muziek, en Theaterwetenschappen, Universiteit Gent,
Sint-Pietersnieuwstraat 41, Technicum Blok 2, B-9000 Gent, Belgium

A. Schneider (ed.), *Studies in Musical Acoustics and Psychoacoustics*,
Current Research in Systematic Musicology 4,
DOI 10.1007/978-3-319-47292-8_9

explores identification of musical intervals realized with sounds from the historic Dumery bells. Findings are evaluated in regard to perception and musical issues.

1 Introduction

Carillons are peculiar musical instruments in regard to both acoustics and perceptual issues. Historic carillons in Europe typically consist of a set of so-called minor third bells (see below) which are often used to play music that, at least in the past two centuries, increasingly made use of the major scale and tonality. In this respect, the sound structure of minor-third bells, on the one hand, and the tonal and harmonic structure of the music played on carillons, on the other, can lead to perceptual discrepancies (see below, Sect. 4).

The main purpose of our article is to study characteristics of the 26 historic Dumery bells that form the fundamental part of the carillon of Bruges. Therefore, the scaling (weights, diameters) of these bells, their spectral structure as well as their tuning to a musical scale has been investigated in detail. We recorded all 47 bells of the carillon in 1997, 1999, and 2004 prior to a restoration in 2010/11, in which the 26 historic Dumery bells (no. 1–26 of the carillon) were cleaned and their tuning was checked (including some minor revisions as was deemed necessary, see [1]. The other 21 bells which had been cast, in 1968, by The Royal Eijsbouts foundry of Asten, the Netherlands, were replaced by new specimens cast by the same foundry. In this article, we will focus on the historic Dumery bells, leaving aside the recent bells cast by Royal Eijsbouts.

For readers not familiar with carillons, we provide some factual and historical background. Also, fundamentals of bell acoustics are outlined with respect to sound generation in bells as well as radiation of sound from bells. Temporal and spectral parameters are given special attention. In addition to sound analyses, some perceptual issues will be addressed since the tuning of bells has often been discussed in conjunction with concepts of pitch perception. A phenomenon that has been investigated for a long time (probably for centuries) is the so-called ‘strike note’ of bells, a sound perceived by listeners immediately after the bell has been excited by a clapper impact. The strike note is considered a decisive factor in bell tuning and in the formation of the pitch or of several pitches arising from the sound of a bell.

Among previous studies on bells and carillons, some have covered aspects of tuning and also perceptual issues (e.g., [2–5]). One study dealt in detail with sound generation in the bell by the clapper impact as well as with the revoicing of carillon bells, which can become necessary due to the wear and tear of both bells and clappers [6]. A recent study [7] addressed reconstruction of the original tuning of a famous historic carillon that had been cast by the bell founder, Willem Witlockx (of Antwerp), in 1730, for the Royal Palace at Mafra, Portugal (see [8]). Of the original 36 bells, only 12 are extant while the carillon had been restored and expanded later on (the latest revision and expansion was made in 1986 by R. Eijsbouts; the carillon

now has 53 bells). Since revoicing and retuning of bells in general involves removal of material from individual bells (even if in small quantity), reconstruction of the original tuning can be quite difficult (see [7]). Luckily, the Dumery bells of Bruges seem to have been left untouched since they were manufactured in the years 1742–1748 [1, 9]. Though the tuning of carillon bells can be affected to some degree by daily use over such a long period, data on bell dimensions and weights (see below) permit to assume that the sounds of the 26 Dumery bells we recorded in the years 1997–2004 still reflect the original tuning closely.

1.1 Some Historical and Factual Background

A carillon is a musical instrument comprising a set of bells tuned to a scale that can be played from a special keyboard or clavier. In addition to the manual, a pedal-board is included in many carillons. Carillons typically cover at least two octaves (23 bells tuned to a chromatic scale) and can have up to 47 bells (four octaves), with the largest instruments comprising 77 bells (Riverside church, New York; see [10], Chap. 11). There are some historic carillons with a smaller ambitus (the world's oldest extant carillon, cast, in 1595/96, by Peter III van den Ghein for the City of Monnickendam, The Netherlands, has 15 carillon bells plus two additional bells used also for a clock). Though carillons nowadays can be found in many places all over the world, much of their history is closely connected with the Low Countries (see [8]). From the historical record we know that carillons including a clavier were introduced in this area, in the 16th century while bell chimes comprising a number of tuned bells that could be activated in some other form by one or several players were in use already in the Middle Ages. In the past, carillons often included bells originally cast as swinging bells to be used in churches or monasteries. As we know from historical sources, the shape of swinging (church) bells has been changed from a more or less cylindrical or beehive design in the Middle Ages to the basically conical structure of bells such as cast by Geert (Gerhardus) de Wou around the year 1500 (for details, see [11]). These bells have a massive sound bow extending the diameter of the nearly cylindrical part (the so-called waist). A bell is closed at the top by a plate that carries the crown with which the bell is fastened to the headstock. At the lower side of the bell wall, the sound bow is continued into the lip or rim which tapers towards to mouth of the bell (see Fig. 1).

The reason for a change in the shape of bells around 1500 apparently was that bell founders attempted to achieve a certain spectral structure in the sound based on the vibration frequencies of several lower partials. The profile of each bell was designed so that the vibration frequencies for the principal partials called hum, prime, tierce, fifth and octave (or nominal) formed ratios like 1:2:2.4:3:4. To be sure, the third partial is a minor third (tierce). The bell following this pattern of partials hence became known as minor third bell (or minor third/octave bell). Once there was a template known for the bell's wall yielding the desired partial structure, one could derive a peel of bells conforming to a musical scale. Basically this was

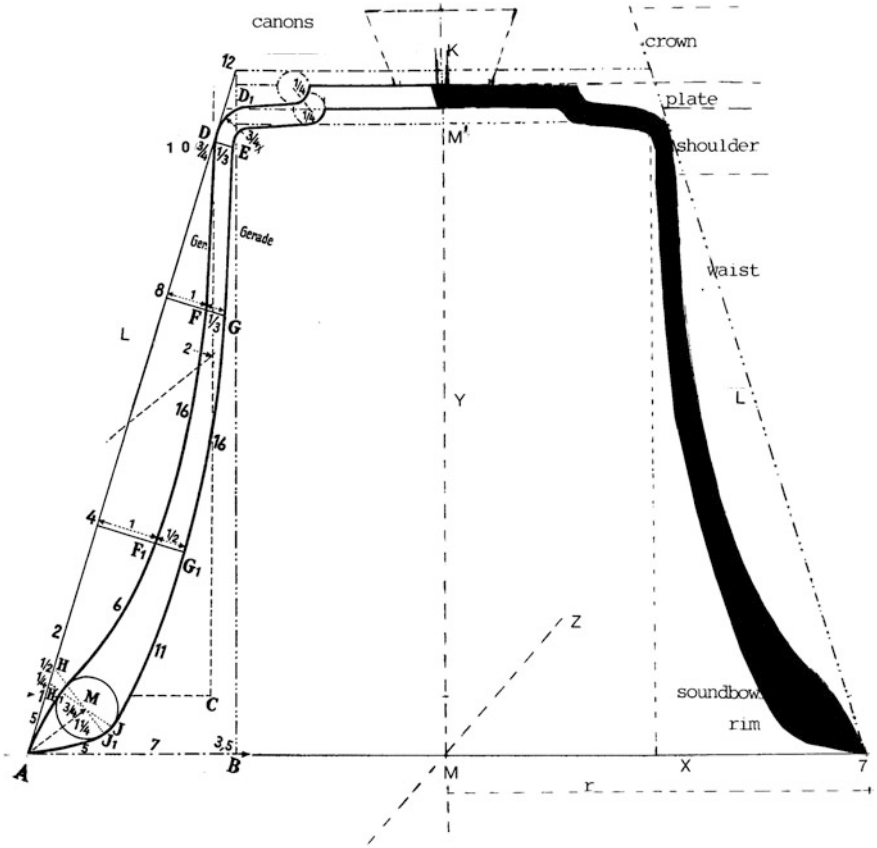


Fig. 1 Bell wall profile, Geert de Wou (adapted from [11])

achieved by scaling (up and down) the overall size of the template as well as the relevant physical parameters (diameter, height, thickness of wall at certain points, weight) associated with the bell shape. However, for reasons of structural stability and also with respect to aspects of sound and tuning there are some limits to such an approach (in particular for small bells; for details, see [11, 12]). Nonetheless, scaling of bells according to a template was and still is a useful approach for designing carillons (see below). One famous bell cast by de Wou, the *Gloriosa* of Erfurt (dating from 1497; see [13, 14]), already shows the pattern of strong spectral components typical of the minor-third/octave bell, that is, the first five major partials exhibit a frequency ratio of close to 1:2:2.4:3:4. In the very large *Gloriosa* bell whose diameter is 257 cm, height ca. 265 cm, weight 11,450 kg, and frequency of the lowest spectral partial is ca. 80 Hz, the actual frequency ratios of the first seven strong components measured from the sound radiated from the bell are close to 1:2.1:2.47:3.04:4.09:5.21:6.1. If the second (and strongest) partial is taken as a reference, the ratios are 0.48:1:1.18:1.45:1.95:2.48:2.91. Hence, this bell and most

other specimen of the minor-third type contain a minor third and a fifth within one octave and a major third as well as a fifth in the next upper octave. The presence of a minor and a major third, which form an interval of (close to) a minor ninth has a profound effect on the tone colour or timbre of the bell sound as well as on the pitch or, rather, the pitches attributed by listeners to such sounds. The partial structure of the minor-third swinging bell served as a model for carillon bells, which evolved in the 16th century and were perfected, in the 17th century, by the famous brothers François and Pieter Hemony who worked in the Netherlands since 1641 (see [8]). The Hemony brothers had met with Jacob van Eyck, an outstanding musician and musical scholar, sometime in the 1640s (cf. [15]). Van Eyck lived in Utrecht where he was involved in the development of carillons in many ways, including tuning of bells ([16], 130ff.). It seems that van Eyck was one of the early scientists who explored the phenomenon of resonance. According to a note contained in the diary of Isaac Beeckman (24th of Sept. 1633), van Eyck had told him that he could hear out some of the partials in bells without touching the bell [17]. Apparently, van Eyck used to sing or whistle a tone in order to excite a resonance in a vibrating body like a bell. He observed some of the partials relative to the strike note (Beeckman mentions the "slach", which in van Eyck's scheme is the octave above the fundamental), most of all, the minor third. Van Eyck seems to have understood that the clarity of the bell's partials as well as the pureness of the pitch perceived from a bell sound, depend on the profile (curvature, thickness at certain points) of the bell's wall. The Hemony brothers put these insights to practical use when they cast bells for carillons which, to this day, are regarded unsurpassed in tonal quality. For all these bells the five lower partial frequencies have a ratio of 1:2:2.4:3:4 (or nearly so).

The City of Bruges had a carillon built, in 1675–80, by Melchior de Haze (of Antwerp) that was housed in the large belfry fronting the market place. However, this carillon was destroyed in a fire, in 1741. As a replacement, a new carillon was ordered from Joris Dumery (Georgius Du Mery) who operated a foundry in Bruges. Dumery promised to deliver "flawless" bells with harmonious sounds (cf. [8] and various documents relating to Dumery's carillon in Bruges in [18]). Dumery's fine carillon originally comprised 45 bells of which 26, cast in the years 1742–48, are extant. To this set, 21 new bells were added, in 1969, cast by Royal Eijsbouts of Asten, The Netherlands, which were designed to match the shape and sound structure of the Dumery bells. The carillon has undergone a revision in 1968/69 which, however, left the original Dumery bells untouched as far as retuning is concerned [9]. As mentioned above, the carillon underwent a complete revision in 2010/11, which included restoration (cleaning, retuning) of the Dumery bells and replacement of the 21 bells from 1969 by newly cast specimen that were adapted more closely to the geometry of the original Dumery bells [1].

1.2 Basic Data Concerning the Dumery Bells

Based on data available from two restoration reports [1, 9] and our own measurements, the following table lists the year of manufacture, the musical notes of the 26 Dumery bells according to their fundamental frequencies f_1, the weight (kg) and the diameter (mm) of each bell. The diameter is always taken at the lip or rim of a bell and hence represents its maximum width at the opening.

The weights, diameters and fundamental frequencies of these bells relate clearly to each other as shown in Figs. 2, 3, and 4. Bell founders knew from experience that bells of identical wall profile and material can be scaled according to proportionality rules which state that two fundamental frequencies relate to each other like the inverse ratio of the diameters of the respective bells, that is $f_1/f_2 = d_2/d_1$. This means that, with identical bell shape, halving the diameter means doubling the fundamental frequency of a bell as is evident from Table 1 and Fig. 2.

Likewise, fundamental frequencies can be related to the weights and diameters of bells where the fundamentals of two bells, f_1 and f_2, relate to their mass (weight in kg) and diameter according to

$$\frac{M_1}{M_2} = \left(\frac{d_1}{d_2}\right)^3 = \left(\frac{f_2}{f_1}\right)^3.$$

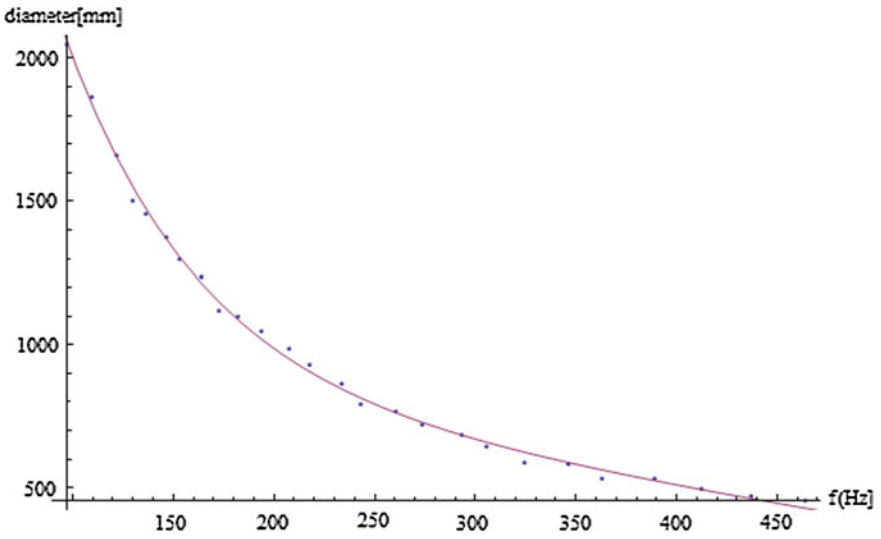


Fig. 2 Diameters and fundamental frequencies of the 26 historic Dumery bells (represented by *dots*)

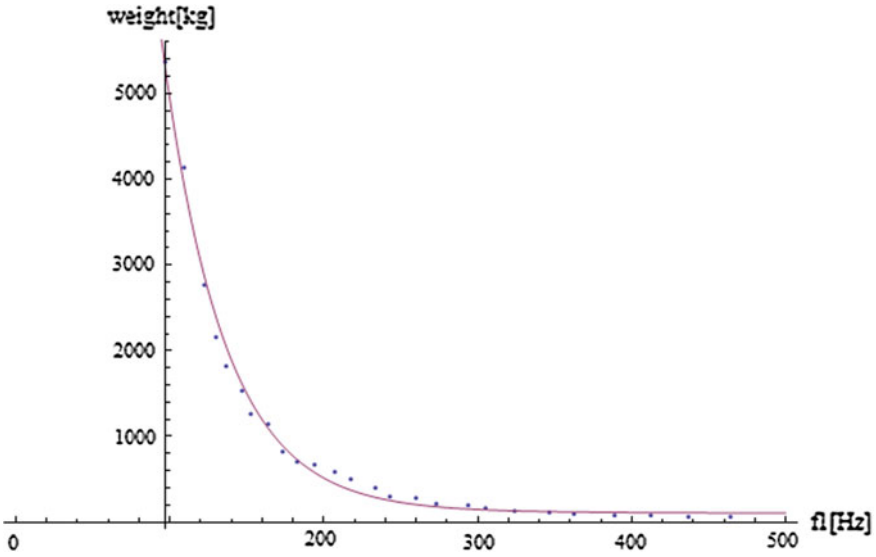


Fig. 3 Weights and fundamental frequencies of 26 historic Dumery bells

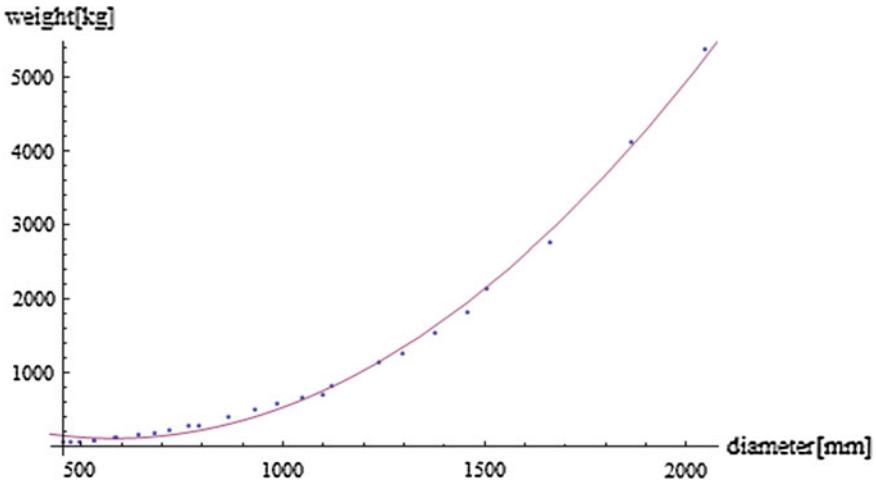


Fig. 4 Weights and diameters of the 26 Dumery carillon bells

This implies that the weight of a bell with a fundamental one octave above that of another bell will be about 1/8 of the bell with the fundamental one octave lower (Table 1 and Fig. 3). Comparing, for example, the G_2 and the G_3 Dumery bells, their weights show a ratio of ca. 8:1. Using the equation above, the cubed ratio of the two diameters yields ca. 7.5 and that of the fundamental frequencies gives 7.86.

Table 1 Basic data, 26 Dumery bells, carillon of Bruges

Bell	Date	Note/Tone	f_1(Hz)	Weight (kg)	Diameter (mm)
1	1744	G_2	97.51	5.378	2047
2	1748	A_2	109.43	4.133	1864
3	1748	B_2	122.05	2.766	1661
4	1743	C_3	129.62	2.153	1503
5	1745	$C\#_3$	136.68	1.825	1457
6	1744	D_3	146.61	1.54	1376
7	1745	Eb_3	153	1.27	1295
8	1744	E_3	164.02	1.145	1236
9	1743	F_3	172.85	830	1120
10	1745	$F\#_3$	182.27	710	1098
11	1745	G_3	193.88	670	1046
12	1745	$G\#_3$	207.51	590	984
13	1745	A_3	217.69	501	930
14	1745	Bb_3	233.84	398	865
15	1743	B_3	243.11	295	793
16	1745	C_4	260.28	281	766
17	1743	$C\#_4$	273.35	222	718
18	1743	D_4	293.56	196	683
19	1745	Eb_4	305.44	165	642
20	1742	E_4	324.18	123	587
21	1743	F_4	346.38	120	583
22	1742	$F\#_4$	362.75	95	533
23	1743	G_4	389.22	81	531
24	1745	$G\#_4$	412.23	77	495
25	1743	A_4	436.65	63	472
26	1743	Bb_4	463.78	61	455

Finally, the weights of bells in a tuned carillon relate to the diameters and the average wall thickness in orderly fashion (see [12, 19]). This can be approximated by the equation $M = c_1 h d^2$, where M is the weight (kg), d is the diameter (m), h is the average wall thickness and c_1 is a constant. The data from the 26 Dumery bells again indicate that halving the diameter and doubling the fundamental frequency coincides with a reduction of the weight to about 1/8 to the bell tuned one octave lower (Fig. 4).

Given the fact that the 26 historic bells (Table 1) show an orderly progression in regard to weight, diameter, and fundamental frequency, one can fit a function to the data for each of the parameters whose graph permits a qualitative assessment of the goodness-of-fit (Figs. 2, 3 and 4). The relatively small deviations of the bell parameter values for the 26 bells from the ideal ratios defined by proportionality rules can be taken as a quality mark indicating that Dumery in fact was able to cast carillon bells at a very high level of craftsmanship and precision of tuning.

2 Basics of Bell Acoustics

Bells belong to the class of musical instruments known as idiophones, literally meaning 'self-sounding' instruments. The basic concept is that a solid (such as a bar, plate, or shell) is used as a vibrating body set to vibration by an impulse affected by means of another body (e.g., a mallet or, in the case of the bell, a metal clapper). Examples of idiophones are xylophones (of which different types are found in Africa and in Southeast Asia), metallophones and gong chimes as are central in Javanese and Balinese gamelan music, and carillons comprising a set of tuned bells. In regard to vibration, a bar of quadratic, rectangular, or circular cross section appears as a relatively simple geometry in particular if the diameter of the cross section is small in relation to the length l of such a bar that can be viewed as a one-dimensional continuum (as in the classical Euler-Bernoulli theory). Likewise, a thin flat plate can be viewed as a two-dimensional structure if the thickness h of the plate is very small in relation to its length and width (in a rectangular plate) or its diameter (in a circular plate). Obviously, in practice one often has to deal with geometries that are more complex. For example, bronze plates such as used for Javanese and Balinese metallophones often exhibit a certain thickness and a trapezoid cross section. Moreover, such plates can be curved to some degree. Due to these factors, in plates of the Balinese *gender* modes of vibration can be identified in addition to the standard pattern known from bars and small flat plates (see [20]). In a swinging or carillon bell, the geometry is even more complex due to the variable diameter and thickness of its wall. The bell wall profile plus the shoulder and plate (with the canons attached to it) make up a compound structure that renders calculation of modes of vibration difficult. Before FEM (Finite element method) and BEM (Boundary element method) modelling became available as standard methodology, calculation of shell vibrational modes usually required simplifications, taking characteristics of standard shell models (such as a cylinder or a spherical shell) as a reference (for in-depth treatment of shell vibration theory, see [21–25], [26], Chap. 7, [27, 28]).

For a simplified model of a bell, one may conceive first of a thin circular plate (2D model) and then of a sphere or hemisphere produced from bending a flat circular plate (3D model). If a circular plate free around its circumference is set to vibration, acoustic figures well-known from the seminal work of Chladni [29] can be observed. Such figures result from the node lines between segments of the plate vibrating in opposite phase. For a circular plate, the number of segments representing modes of vibration can be ordered according to nodal meridians (m) and nodal circles (n). Such patterns can be observed (with certain modifications) also in square as well as in rectangular plates (see [19]). Vibration where segments of a plate move inward and outward in general results from flexural or bending waves, which are acoustically the most effective in regard to sound radiation (see below). However, in bars and plates also longitudinal motion occurs inside the structure, as well as torsional vibration due to tangential motion. Since elastic solids exhibit

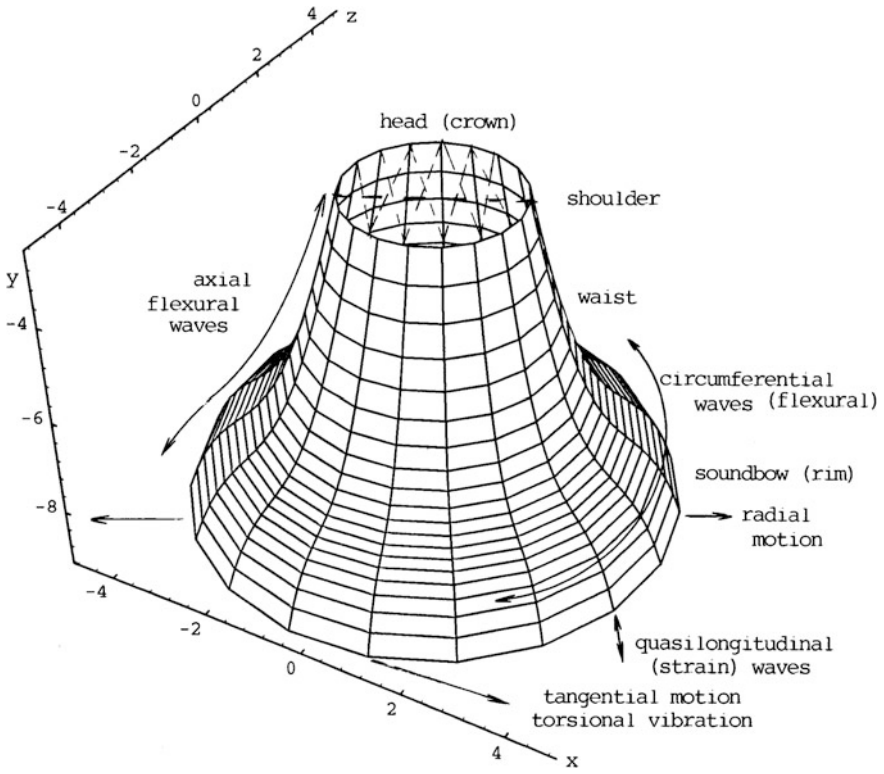


Fig. 5 Types of vibration in a bell-like structure

stress and strain when forced to vibrate, a special type of strain wave labelled quasi-longitudinal is also found.

Taking a sphere or hemisphere as a model, it is straightforward to assume that such a structure, when struck with a mallet at a point on the wall near the edge, may allow flexural motion of its wall around its circumference as well as possibly in axial direction. Furthermore, tangential motion seems feasible. In addition, a structure such as a plate or a hemisphere can exhibit quasi-longitudinal motion within the thickness of a (flat or curved) plate. The respective types of vibration are usually labelled flexural or bending, torsional, and quasi-longitudinal (due to stress/strain), respectively. Taking a structure closer to a real bell as model, the types of motion can be indicated as in Fig. 5.

Wave propagation for flexural motion in such a structure differs from propagation in fluids (such as air) in that shear forces between molecules occur. Consequently, for elastic solids one has to take into account bending stiffness, rotatory inertia and other parameters relating to the geometry as well as to material properties of vibrating structures (see [23, 26–28], [30], Chap. 7, [31]). Wave propagation in solids differs from fluids in that the phase velocity for bending waves

is dependent on the frequency of its components. Hence, each component in a wavepacket travels with a particular phase velocity, c_B. While wave propagation in air is constant at $c = \lambda f$ and $\lambda = c/f$, phase velocity for propagation of bending waves in a thin plate of unlimited extension and free around its circumference approximately is $c_B \sim \sqrt{f}$. Phase velocity thus grows in proportion to the root of each frequency component, and is inversely proportional to the wavelength for which $\lambda_B \sim 1/\sqrt{f}$. More precisely, phase velocity in a homogeneous plate is

$$C_B = \sqrt{\omega}\sqrt[4]{B'/m'},$$

where $\omega = 2\pi f$, B' is the bending stiffness, and $m' = \rho h$ is the mass per unit area of the plate. Bending stiffness in a homogeneous plate can be calculated as

$$B' = \frac{E}{1-\mu^2} I',$$

where E is Young's modulus, μ is the Poisson ratio (the value is 0.3), and $I' = h^3/12$ is the geometrical or area moment of inertia. Doubling the thickness h of a plate of given dimensions not only doubles its mass but raises B' by a factor of eight. If thickness h is of significance, bending stiffness has to be calculated as

$$B' = \frac{Eh^3}{12(1-\mu^2)} \left[\mathrm{N\,mm^{-1}}\right];$$

phase velocity then becomes $C_B = \sqrt[4]{\frac{Eh^2}{12(1-\mu^2)\rho}}\sqrt{\omega}$.

In a dispersive medium like a plate, besides phase velocity and group velocity also the wavenumber k_B for bending waves is dependent on frequency:

$$k_B = \sqrt[4]{\omega^2 \frac{m'}{B'}}$$

The equation for bending waves in plates can be written as

$\Delta\Delta\xi + \frac{m'}{B}\frac{\partial^2 \xi}{\partial t^2} = 0$ where $\Delta\Delta\xi \equiv \nabla^4 \xi$; the 4th differential parameter (G. Lamé) is defined as

$$\Delta\Delta\xi = \frac{\partial^4 \xi}{\partial x^4} + 2\frac{\partial^4 \xi}{\partial x^2 \partial y^2} + \frac{\partial^4 \xi}{\partial y^4}$$

The equation of motion for a circular plate with free edge and thickness h can be written in plane polar coordinates (r, φ)

$$\Delta\xi = \frac{\partial^2 \xi}{\partial r^2} + \frac{1}{r}\frac{\partial \xi}{\partial r} + \frac{1}{r^2}\frac{\partial^2 \xi}{\partial \varphi^2}$$

where ξ is the displacement of particles and r is the radius of the plate. Solutions are found from linear combinations of Bessel functions $J_m(x)$ and hyperbolic Bessel functions $I_m(x)$. Frequency ratios for bending waves (relative to $m = 2$, $n = 0$ taken as 1.00) in a circular plate are inharmonic (for both $\mu = 0.25$ and $\mu = 0.3$) though some approximately harmonic ratios can also be found.

For a shell of constant radius and thickness, bending stiffness can be calculated as in thick plates; strain stiffness D of a shell is calculated like $D = \frac{Eh}{1-\mu^2}$ [N mm^{-1}]; from B' and D a factor for shells of radius r and thickness h can be calculated, which is

$$\beta = \frac{B'}{D}\frac{1}{r^2} = \frac{h^2}{12r^2}$$

The equation of motion for bending waves in shallow spherical shells can be derived from the equation of motion for flat plates by inserting an additional term (see [19] Chap. 3), which is $\nabla^2 H/R$. Here H is Airy's stress function and R is the radius of the shell's curvature. The example demonstrates that shells of simple geometry can be viewed as derived from flat or curved plates (see also Junger and Feit [26], Chap. 7). Also, several so-called 'membrane' theories of shells (see [21], Chap. II–IV, [23], § 4.3; the name was chosen for some analogies to vibration theory of membranes) assume a very thin, elastic shell in order to simplify parameters relevant for vibration. However, for bending waves in shells of a more complex geometry including thick walls (that is, h is significant relative to the length l of a cylinder or cone), displacement kinematics needs to include shear forces and other factors (see [21], Chap. V, VI, [22], [23], § 8.34). In sum, calculation of modal frequencies in bells is complicated because parameters such as variable diameter, variable wall thickness and, hence, variable bending stiffness, have to be taken into account (see [24, 25]).

From the equations stated above it is evident that propagation of bending waves in solids depends on the geometry and the material parameters of the structure that is vibrating free from outside forces once the impulse force $F(t)$ necessary to excite normal modes of vibration has been applied. Frequency dispersion in solids means that, for bending waves (flexural motion), eigenfrequencies do not form harmonic series but must be determined for certain geometries (such as rectangular or circular plates, cylindrical or conical shells) according to boundary conditions (the most relevant conditions concern free, simply supported or clamped edges) and material parameters (see [27, 31]). Spectra of bars, plates and shells undergoing flexural vibration are essentially inharmonic (for a detailed account including empirical data, see [32]).

2.1 Material and Shape

Bells are made of some alloy (such as bronze, brass, iron or steel) which has characteristic properties relevant for vibration behaviour. In modern Western bells, bronze in general contains about 78–80 % Cu and 20–22 % Sn (plus some small parts of other metals such as Pb, which can be found in a more significant percentage in historic bells). For bell bronze, $E = 9650\ \mathrm{kp/mm^2} = 94.14\ \mathrm{GPa}$, $\mu = 0.3$ and density $\rho = 8400\ \mathrm{kg/m^3}$ (for more technical data, see [33]). As to the shape of bells, one can find more or less cylindrical or 'beehive' specimen (as in East Asia) as well as basically conical structures in most Western swinging and carillon bells, which can be derived from a trapezoid representing the cross section of a cone truncated at the peak. Many bell founders used such a frame for drawing a bell's wall profile by means of compasses (see Fig. 1). The wall viewed relative to a symmetry axis located right in the middle of the bell obeys to the principles of a rotational shell with radius r and height l. The basic cylindrical or conical shell in the case of common bells hanging from a support is closed at one end by a stiff and relatively thick plate which, in the case of western swinging and carillon bells, carries canons used to fastening the bell to a headstock (which in turn can be part of a bell frame).

Due to the axisymmetrical shape of the bell mantle, mass distribution ideally is even around the middle axis. However, the geometry of most of the historical swinging bells found in many countries in Europe as well as bells that are used in carillons is more complex in that the radius r, and hence the diameter d of the bell vary along its middle axis $M - M'$(see Fig. 1) as does the thickness h of the bell's wall. Thereby, also bending stiffness B varies with respect to the bell profile. At the so-called soundbow (or ring), stiffness is large consequent to the thickness of the bell's wall at this point whereas at the so-called shoulder of the bell, stiffness again is considerable yet has to be attributed to the shape itself, namely the circular plate adjacent to the shoulder which closes the bell. The plate carries the canons (not detailed in Fig. 1) necessary to fasten a hanging bell to a support. Furthermore, thickness of the wall and stiffness are factors that bear to the internal damping (attributed to stress in the material) of the bell. Because of these features, the vibration theory for this type of compound shell is quite complex (for vibration theory of shells, see [21–23, 25–28]).

2.2 Excitation of Normal Modes and Radiation of Sound

Western swinging and carillon bells are set to vibration by applying a force via a clapper. In carillon bells, the clapper is activated by a player from a clavier from which a transmission leads to the clapper inside the bell. Alternatively, vibration in carillon bells can be excited by mallets in case carillon bells are connected to a clock-work which activates the mallets. While the clapper usually hits the inner

wall at the soundbow near the rim, the mallet excites the bell from the outside (at about the same region on the bell's wall). For the clapper hitting the wall, the contact time is quite short at, in general, 0.7–1.5 ms depending on the force applied (and, thus, on the velocity of the clapper motion) as well as on the mass and material of the clapper ball (for data, see [6, 24, 34]). The clapper or hammer transmits a force $F(t)$ via the impulse; the magnitude of the force depends on the impedance, Z, at the bell's surface (in regard to flexural vibrations which are of foremost concern). In earlier experiments it was found that, roughly, the mass of the clapper and the contact time are proportional, that is, doubling the mass of a clapper means approximately doubling the contact time with the bell. This in turn leads to diminishing amplitudes of higher partials in the sound spectrum (cf. Grützmacher et al. [24], 41–43). Impact dynamics in carillon bells investigated in detail [6] revealed that, for an impedance of a bell of $Z = 3 \times 10^4$ kg s^{-1} (a realistic value for a bell of given dimensions), the clapper comes to rest against the bell's wall, and is pushed back immediately by a returning vibromotive pulse. Hence, a very short contact time ensures maximum energy transfer to the bell which, in turn, means that very many eigenmodes are elicited. Since the contact relates closely to the shape of the impulse (duration, height), it is also of influence on spectral energy distribution; short contacts account for 'brighter' sounds because many higher modes are elicited. It has been suggested that re-voicing of carillon bells (whereby the original curvature at the site of the clapper/hammer impact is restored) is suited (a) to increase the impact duration so that the sound becomes more 'mellow' (and less inharmonious), and (b) to make the impact time more dependent on impact velocity whereby 'strong' playing (Dutch: *sterke slag*) makes notes both louder and brighter (the same effect is observed when playing a piano at various dynamic levels). In fact, carillonists can influence the force transmitted by an impulse to a bell by their individual way of playing and can thereby control the dynamics and spectral structure of bell sounds to some degree.

The sound in a bell is produced by vibrations of the bell's wall (including the top plate to some extent) in a pattern of eigenmodes at certain eigenfrequencies. While several lower normal modes can be identified quite easily by detecting the number of nodal meridians and nodal circles (see below), it may be arduous to investigate higher modes due to the small areas involved as vibrating in opposite phase and also because certain modes have nearly identical eigenfrequencies. Due to the bell's axial (rotational) symmetry, modes of flexural vibrations with $m > 2$ are found in doublets (called *Zwillingstöne* in German terminology). This means that such modes occur in nearly degenerate pairs (cf. [35] where the two mode frequencies differ but little. In vibrating systems, two (or more) eigenfunctions and eigenmodes that produce the same eigenfrequency are called degenerate. This condition is often met in quadratic membranes and plates where ω_{mn} can be the same as ω_{nm} (e.g., $\omega_{21} = \omega_{12}$; see [36], 176). However, for the so-called 'breathing mode' of shells ($m = 0$), only one natural frequency obtains (cf. [26], Chap. 7), that is, this mode is a singlet. In case the rotational symmetry of the bell around its middle axis would be perfect, the meridians of one member of the degenerate pair would be found lying exactly on the vibration antinodes of the other.

Historically, the development of the theory of vibration for bells did stem from calculations Leonard Euler provided for vibrating rings. He viewed a bell as a series of *annuli elementares* (simple rings, see [37]). One in fact may break down the bell's mantle into several ring-like segments. For Asian bells close in shape to a 'beehive', conical ring elements were found suited to FEM modelling (cf. Chung and Lee [38]).

Lord Rayleigh [28], Vol. 1, §§ 232–235, on the basis of a cylindrical shell taken as a curved plate, gave a detailed account of the modes of vibration found in a typical bell with (almost perfect) rotational symmetry. He considered flexural vibrations around the bell's circumference (the zero points of which result in nodal meridians at equal distances), and along the bell's axis (the zeros of which result in nodal circles). Nodal meridians and nodal circles divide the surface into segments. The number of nodal meridians and nodal circles defines the number of segments which vibrate opposite in phase to each other (see [24]). Since the corpus of a typical western bell consists of a massive ring near the mouth plus a more or less cylindrical shell (i.e., the waist) added to it (see Figs. 1 and 2), certain modes of vibration appear to be 'ring driven' while others are regarded as being 'shell driven'. It should be mentioned that, in the terminology used by Charnley and Perrin [39] in regard to bells, axial motion is labelled 'meridian-tangential', and tangential is specified as 'ring-tangential'.

Modes of vibration in bells have been analyzed with various methods, and have been described and classified in great detail (cf. [24, 35, 40], [19], Chap. 21.1, [41–45]).

As with other vibrating systems, one has to distinguish between inextensional and extensional modes of vibration (see [28], § 232), the latter involving stretching of the bell's corpus whereas for inextensional modes a neutral ring for each radial plane can be assumed. Though flexural vibrations are most prominent in bells, torsional vibration was also observed [39]. Torsional ('twisting') vibration is well known from rods. In such structures as well as in plates, another type of vibration is found usually labelled 'quasi-longitudinal' (see [30], 78ff.). Quasi-longitudinal waves involve strain/stress of the material along the longitudinal axis as well as contraction of the cross section. In a bar or rod of length l (x-axis) particles thereby are displaced also in the direction of the y- and the z-axis, respectively (in a plate, this motion is almost restricted to the z-dimension); therefore, phase velocity c_D for such waves is considerably smaller than that of pure longitudinal waves. It seems reasonable that quasi-longitudinal (strain) waves might occur also in the wall of bells, in particular in the waist where the wall is relatively thin. There are some indications for such types of vibration in historical bells; in one instance, a mode of vibration was reported having a frequency close to that of the nominal, and being strong enough to interfere with that partial (see [41], p. 2004).

Among the methods that have been employed to investigate normal modes of vibration, and to visualize patterns of vibration in bells, are time-averaged hologram interferograms [44], finite elements (FEM, see [13, 42, 46, 47]), and modal analysis (MA, see [7, 40]). A special study on the mode structure formulated as a dynamic

contact problem of bell and clapper was undertaken by Lau et al. [34] with FEM methodology.

Investigations of bells are mostly confined to types of flexural vibration because these cause motion normal to the bell's surface. Since the walls of the bell couple directly to the sound field, motion normal to the surface will radiate most of the sound that becomes audible. Since efficient radiation of bending waves from the vibrating surface of a bell requires that the wave speed C_B must be at least equal to, or greater than, the sound speed in air (340 m/s; see [30], 457ff.), it follows that modes of vibration of higher order can be more prominent in the spectrum than those of lower order.

Radiation of sound from the bell's surface is quite complex due to the different size parts of the wall cover for various modes of vibration, and also due to the curvature of the wall. Consequently, there are different directivity patterns for different normal modes ([47] and additional material presented in a lecture). As to the decay of individual partials, there are several types of damping to be considered. First, damping is small within bell bronze (with a very low damping factor $\delta \approx 0.0004$) in case the material is homogeneous in molecular particle structure and of little porosity (see [33]). Structural damping resulting from the shape of the vibrating body plays a role at the bell's shoulder where the more or less cylindrical or conical part joins the plate at an angle close to 90° (see Fig. 1). Further, there is viscous damping at the boundary between bell and surrounding air as well as acoustical damping within the sound field (as to aspects of sound radiation from bells including damping, see [5, 24], [19], Chap. 21.11, [44, 47). Internal as well as acoustical damping for at least some of the normal modes in bells seems to be rather small since the decay time in particular of the hum can be very long. For the *Gloriosa* of Erfurt (a huge bell, see above), the sound reportedly was audible for 310 s ([14], 113; in 1985, the *Gloriosa* underwent a major repair, which extended the decay time to ca. 370 s). As a rule of thumb, for large bells a decay time (60 dB from initial level) of the partials dominant in the sound (in particular the prime and the minor third) of $t_d > 10$ s seems reasonable. The hum, though usually relatively weak in amplitude, would be audible much longer..

Note that in carillon bells, a very long decay of partials is not necessarily useful since it could hamper perception of melodic lines (and particularly so when a polyphonic piece of music is played). The overall sound intensity measured from bell no. 2 as radiated in one direction (the site of the microphone used in the recording) shows a decay of more than 40 dB in less than 4 s from maximum (Fig. 6).

The normal modes of vibration (m, n) typical for octave minor third bells, that is, for bells which have the interval of a perfect (or nearly so) octave between hum and prime as well as between prime and nominal, and a strong minor third above the prime (see [41, 48]) are listed in Table 2.

The number of nodal meridians is either given as full meridians (extending over the top of the bell to the opposite side of the wall) or as half meridians ($2m$). Full meridians in Table 2 are given in brackets. In most cases, the partials that have from 4 to 26 nodal meridians but only one nodal circle (or none, as is the case with the

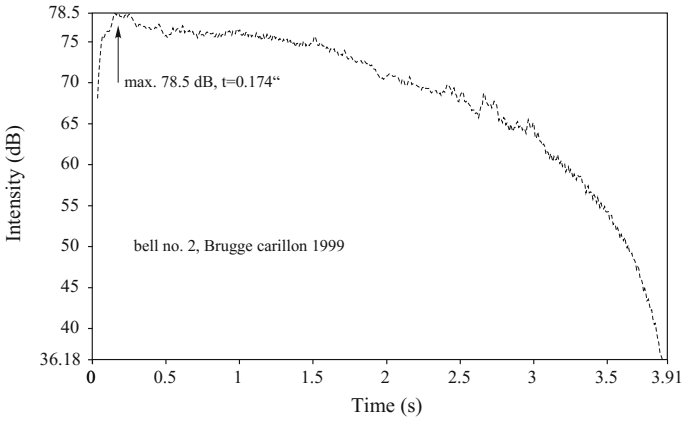


Fig. 6 Sound intensity decay over time, bell no. 2, Brugge carillon

Table 2 Scheme of bell partials and respective modes of vibration

Partial	Frequency ratio to prime	Mode (nodal meridians *m*; nodal circles *n*)
Hum	0.5	4, 0 [2, 0]
Prime [fundamental]	1	4, 1 [2, 1]
Tierce [minor third]	1.2	6, 1 [3, 1]
Quint [fifth]	1.5	6, 1 [3, 1]
Nominal [octave]	2	8, 1 [4, 1]
Tenth [major third]	2.52	8, 1 [4, 1]
Twelfth [fifth]	3	10, 1 [5, 1]
Thirteenth [major sixth]	3.36	10, 1 [5, 1]
Double octave	4	12, 1 [6, 1]
Upper fourth	5.33	14, 1 [7, 1]
Upper major sixth	6.73	16, 1 [8, 1]
Triple octave	8	18, 1 [9, 1]
Minor third	9.5	20, 1 [10, 1]
Fourth	10.68	22, 1 [11, 1]
Fifth	12	24, 1 [12, 1]
Major sixth	13.45	26, 1 [13, 1]

hum) will form the strongest peaks in spectra obtained from actual bell sounds recorded in the free field. The phenomenon that there are pairs of modes that have identical numbers of meridians and circles yet represent different partials (for example, 6, 1 [3, 1], comprising the minor third and the fifth above the fundamental), stems from the fact that the position of the nodal circle on the bell's surface can vary considerably. The nodal circle for the minor third (6,1) will be found on

the waist while the nodal circle for the fifth (6,1) is located near the soundbow of the bell (cf. [19], Chap. 21.1).

Even though the vibrational modes listed in Table 2 in many cases will be the most prominent ones, other modes can be found. For example, the eleventh (ratio to the prime/fundamental: 2.67) has six half-meridians that are equivalent to three full meridians conceived of as extending over the crown of the bell, and two nodal circles ([13, 14]). In measurements based on acoustical excitation of an English church bell, no less than 134 partials representing modes of vibration have been found in the frequency range from 292 Hz (hum note) up to 9300 Hz [42]. Of course, it is not possible to relate all these modes to partials that constitute sections of harmonic series such as will be found in Table 1. In fact, actual bell sounds often contain many more inharmonic components than just the minor third which in many bells is strong in amplitude, and therefore is characteristic of most of our church and carillon bells (see Table 3). The minor third spectral component can interfere with the partial of the major third (or tenth) that is found one octave higher. The spectral composition of minor-third bells, which is inharmonic to some extent, has at times been found unsuited to rendering musical pieces written in a major key. In fact, performance of a piece in a major key played on a typical minor-third carillon may appear ambivalent in regard to perception and musical composition. This issue played a role in the development of major third bells that, to be sure, have a much different geometry and shape of the wall's profile (see [49]).

In case the rotational symmetry of the bell around its middle axis would be perfect, the meridians of one member of a degenerate pair would match the vibration antinodes of the other. Since especially historical swinging bells rarely have been cast to result in perfect mass symmetry, and may exhibit both variations in the thickness along the wall as well as deviations from a perfect ring with respect to the cross section, the two members of a pair have different vibration frequencies. In the spectrum of the bell sound one therefore quite often finds twin peaks representing the two members of a (nearly) degenerate pair. The distance of the peaks increases with the amount of deviation from perfect axial symmetry. Typically, the frequency difference of the two members of such a pair is from less than 1 Hz to a few Hertz (in cents, the difference often is less than 100 cents and sometimes even < 50 cents). Perceptually, the effect can range from a slight shimmering of spectral components (comparable to the 'chorus' effect applied to electric guitars) to audible amplitude modulation (AM) that will be registered as beats or roughness. In case the effect is pronounced, it usually is labelled warble (cf. [50, 35]). There are indications that bell founders of the past deliberately may have allowed small deviations from symmetry; slightly eccentric shapes where degenerate pairs are separated into two independent components possibly were employed to reduce warble. The effect of warble as a source of amplitude modulation is shown in Fig. 7 for bell no. 18 of the Dumery carillon. AM is significant and fairly regular relative to the average decay of the sound level (indicated by the dashed line in Fig. 7).

Though spectra of bars, plates and shells are essentially inharmonic, the profile of the typical minor third bell yields spectral components many of which can be assigned to several harmonic series that are, however, incomplete. Also, there are

Table 3 Bruges, bell no. 1, partial structure, main components 0–3 kHz

No.	*f*[Hz]	A [dB]	ratio f_n/f_1 [ss]	*f*[Hz]	A [dB]	Ratio f_n/f_1 [ir]	Partial name
1	97.47	53.0	1	97.51	68.3	1	Hum
2				149.95	40.3	1.54	
3	195.69	66.1	2.0	195.74	71.3	2.0	Prime
4	233.48	64.8	2.395	233.49	68.3	2.395	Tierce (minor third)
5	294.65	44.7	3.023	294.94	28.4	3.025	Quint (fifth)
6	391.37	67.7	4.015	391.34	42.2	4.013	Nominal (octave)
7	488.91	40.4	5.016				Tenth (major third)
8	493.50	47.8	5.063	493.54	39.1	5.061	Tenth (major third)
9	515.27	47.6	5.268	515.29	44.2	5.284	
10	525.59	54.7	5.392	526.40	35.1	5.398	
11a	574.09	44.0	5.89				
11b	577.37	46.9	5.923	576.98	23.8	5.917	
12	590.02	60.4	6.053	590.15	44.9	6.052	Twelfth (fifth)
13	628.55	44.1	6.45	628.70	30.8	6.448	Thirteenth (sixth)
14	683.79	34.3	7.015				
15a	819.44	50.1	8.41	819.80	51.9	8.407	‘Double octave’?
15b	819.79	50.5	8.41				
16	1073.67	54.1	11.015	1073.63	48.2	11.01	Upper fourth
17a	1091.47	42.0	11.2				
17b	1091.82	42.0	11.2				
18	1197.84	30.6	12.29	1197.96	28.2	12.28	
19	1345.68	46.7	13.8	1345.94	44.8	13.8	Upper major sixth
20	1630.94	35.5	16.73	(a) 1631.14	32.5	16.73	Triple octave
				(b) 1631.54	34.5	16.73	
21	1925.31	38.0	19.75	1925.14	36.1	19.743	Minor third
22	2223.31	27.7	22.81	2223.69	39.0	22.805	Fourth
23	2524.63	19.8	25.9	2524.91	38.8	25.893	Fifth
24	2826.01	28.2	29.0	2826.04	39.8	28.98	

components that have inharmonic frequency relations both with the prime and the hum (either can be taken as fundamental), and among each other. Furthermore, because of the profile necessary to produce the minor third, the octaves even in many swinging bells designed as so-called octave bells are hardly perfect, and rather tend to show some characteristic deviations (cf. [48], also data in [51]).

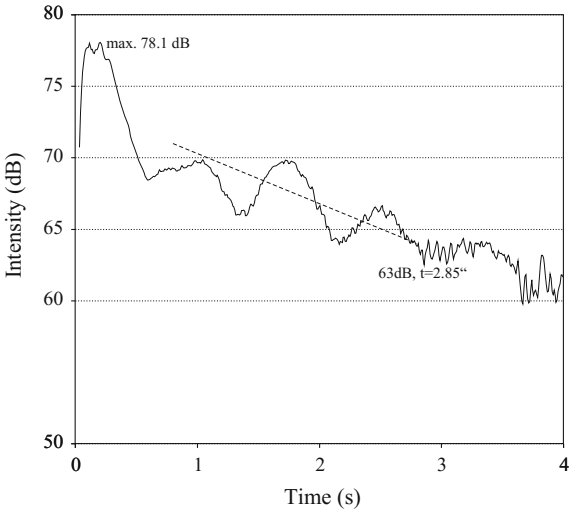


Fig. 7 AM in the sound of bell no. 18; average decay of level (*dashed line*)

3 Inner Harmony and Tuning

As stated above, bell profiles since about 1500 were designed to produce the first five major partials in frequency ratios as close as possible to 1:2:2.4:3:4, and with partials above this series matching the scheme listed in Table 2 in good agreement. In practice, however, one will always find deviations from these ideal frequency ratios, and one may also encounter spectral components interspersed, as so-called 'mixture partials', between the 'principal partials' hum, prime, tierce, etc. In order to evaluate sound characteristics of individual swinging and carillon bells, it is always useful to study those individual modes of vibration that can be excited in the bell (what can be done with various methods, see Grützmacher et al. [24], [41, 42]) as well as to identify partials and other components in the sound radiated from a peculiar bell. Table 3 lists frequencies from sound spectral analysis (recordings in 1999 and 2000) as well as from impulse response measurements (recorded in 2014) taken from bell no. 1 of the carillon. Frequency ratios are given for a comparison of the partial structure before and after restoration. The amplitudes also included in this table should be taken as a relative measure only (since depending on the site of recording); however, the readings indicate the relative strength of partials within each set of data (sound spectrum = ss; impulse response = ir).

Data from both measurements in general agree very well (indicating that retuning of this bell, in 2010, has been very slight, if detectable at all). Some of the partials (see nos. 1, 3, 4, 5, 6, 7, 12) are very close to harmonic ratios while some other deviate markedly (e.g., the double octave, which is clearly stretched), and thus create inharmonicity in the spectrum. Moreover, there are some doublets (pairs of closely spaced frequencies such as 11a/11b) which introduce amplitude modulation besides the 'chorus effect' they exert on spectral pitches. In the data representing the

impulse response only one such doublet is included (at the triple octave). Detecting doublets (which can occur for normal modes $m > 2$) with accelerometers depends largely on where these are put around the circumference of the bell relative to the point of excitation. However, doublets are contained in the sound radiated from the bell where their frequencies can be precisely determined by means of spectral analysis using FFT of sufficient window size plus suitable peak interpolation (if needed). Small differences in frequency between a nearly degenerate pair of eigenmodes indicate that there is a slight deviation in the bell's geometry and mass distribution from a perfect axisymmetrical pattern (see above).

The sound spectra of all 26 Dumery bells were analyzed in detail to identify the range of partials within a band of 0–3 kHz for the larger bells, and up to 5 or 6 kHz for the smaller bells. Taking those partials into account that carry sufficient energy to contribute to the overall sound within the first 1–2 s after excitation of the bell, there are in general some 30–40 spectral components to be considered. See, for example, the spectrum of bell no. 9 (Fig. 8) where the partials up to 5.4 kHz are displayed. Partials 1–5 (hum, prime, tierce, fifth, nominal) are marked with nos. 1–5. The prime (no. 2) in this sound is the strongest component while the fifth (no. 4) is rather weak. One can see several strong partials located between the nominal (octave, no. 5) and the double octave (marked DO).

Among these components, there are often doublets arising from nearly degenerate pairs of eigenmodes. For example, in bell no. 2 (Fig. 9) the hum note contains two closely spaced frequency components (f_{1a} = 109.21 Hz, A_{1a} = 64.1 dB; f_{1b} = 109.45 Hz, A_{1b} = 66.7 dB).

Since the so-called inner harmony of a minor-third bell rests on the first five principal partials whose frequency ratios should be as close as possible to 1:2:2.4:3:4, one can check the overall tuning of a set of carillons by plotting the log frequencies of these five partials as a function of the scale (expressed in HT = half tones). If the bells of the carillon are tuned in the same pattern of just or nearly just intervals, all frequencies representing the same partial would fall on a straight line. Figure 10 shows the frequency trajectories of 26 Dumery bells for the five principal partials hum, prime, tierce, fifth and nominal (or octave). One can see that most of the bell partials in fact match these trajectories fairly well though a few deviations from the template can be observed.

The hum, the prime, the tierce and the nominal (octave) are almost perfect while there is some variation in the trajectory representing the fifth. One should remember, however, that the fifth in general is weaker in amplitude (as well as in SPL) than the often very strong partials prime, tierce, and nominal. Hence, in regard to perception deviations from ideal frequency ratios for the fifth will not count as much as would deviations in these strong partials. To be sure, there is another partial that deviates from ideal frequency ratios in a more or less systematic way: the double octave typically has a frequency ratio of about 8.4 or 8.5 to the hum. For technical reasons (cf. [12, 50, 41]) bell tuning has to seek a compromise between the tuning of the first five partials, on the one hand, and the double octave as well as probably some more higher partials, on the other. To have the frequencies of the low principal partials 1–5 fall into place, the double octave must be stretched. For

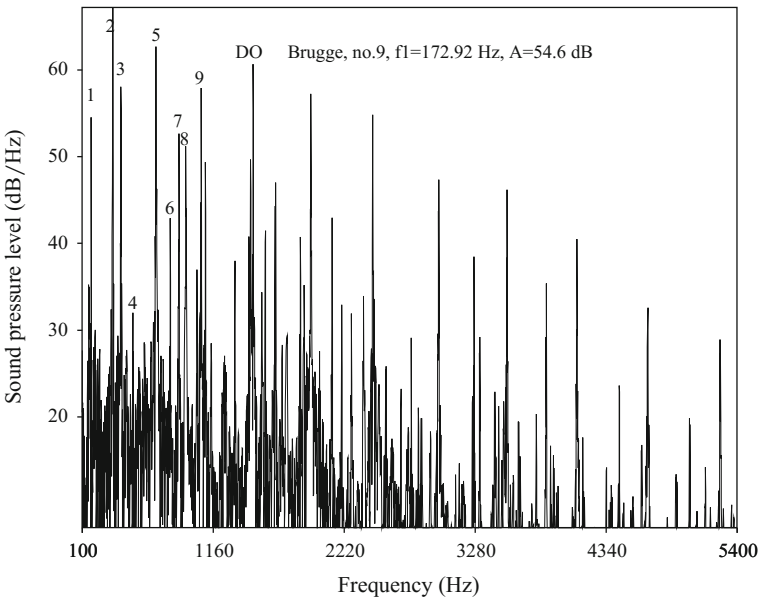


Fig. 8 Bruges, spectrum of bell no. 9. Partials 1–9 marked with nos. 1–9; *DO* double octave

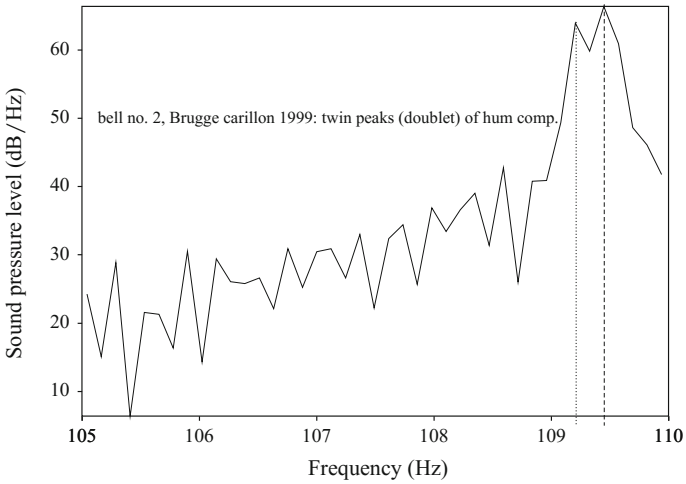


Fig. 9 Doublet, hum partial, two frequency components, bell no. 2

example, in bell no. 2 the frequency ratios are like 1:1.999:4.02:8.41 for the hum, the prime, the nominal, and the double octave. Stretching the double octave coincides with stretching the frequencies of other partials in higher octaves. Note that the double octave usually is a strong partial which, moreover, is in a frequency

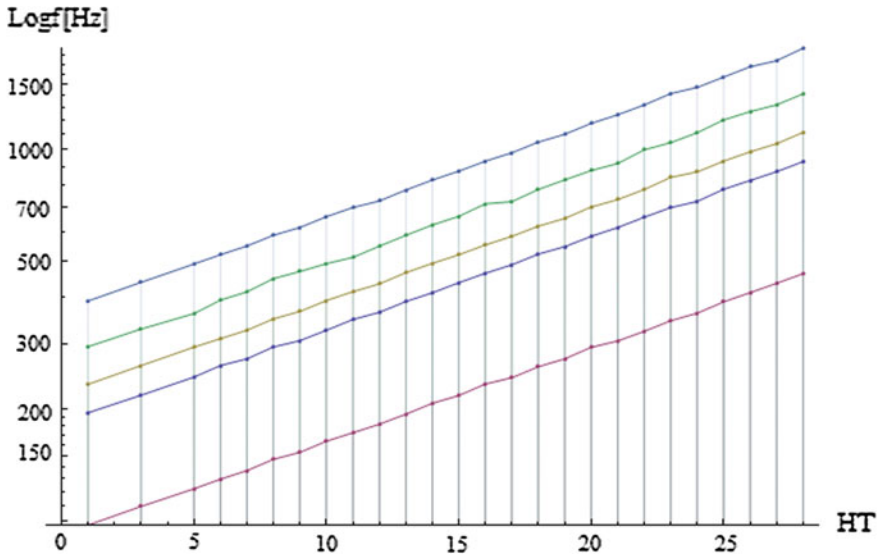


Fig. 10 Trajectories for the principal partials hum, prime, tierce, fifth and nominal (octave), 26 Dumery bells. The abscissa is ordered according to half tone intervals (no. 1 = G_2 bell, no. 3 = A_2 bell, no. 5 = B_2 bell, no. 6 = C_3 bell etc.)

region where the ear is more sensitive than in the low frequency region where the hum and (in large bells) even the prime are found. The stretching of the double octave and the other partials in its vicinity is a source of inharmonicity in the sound. It adds both to the ambiguity of pitch perception and to the sensation of a sometimes shimmering timbre.

Taking the first strong partial (the hum) as the fundamental frequency f_1 of each bell, the tuning can be based on this objective acoustic parameter. If the intervals between the fundamental frequencies are transformed into cents, the tuning can be viewed in musical terms. Table 4 contains all intervals for 26 Dumery bells in a matrix.

It has been speculated at times that the Dumery carillon might have been tuned to a meantone temperament, of which quarter-comma was the most common type (see [52]). A scale of twelve tones would consist of the following intervals:

Tone	1	2	3	4	5	6	7	8	9	10	11	12
Cents	0	75.5	193	310.5	386	503.5	579	696.5	772	889.5	1007	1082.5

Characteristic of this tuning is that it features the pure major third (frequency ratio 5/4) as the basic structural interval while so-called 'Pythagorean' tuning is based on a progression in pure fifths (ratio 3/2). Quarter-comma meantone tuning (see also [53]) means that the comma of nearly 22 cent marking the difference

Table 4 Tuning matrix (cents) of 26 Dumery bells based on f_1 measurements

	1 G (G2)	2 A (A2)	3 H (B2)	4 c (C3)	5 cis (C#3)	6 d (D3)	7 es (Eb3)	8 e (E3)	9 f (F3)	10 fis (F#3)	11 g (G3)	12 gis (G#3)	13 a (A3)
1 G (G2)	0												
2 A (A2)	198.6	0											
3 H (B2)	385.5	188.9	0										
4 c (C3)	489.7	293.1	104.2	0									
5 cis (C#3)	581.6	385.0	196.1	91.9	0								
6 d (D3)	703.0	506.35	317.45	213.25	121.4	0							
7 es (Eb3)	776.9	580.23	391.34	287.12	195.25	73.9	0						
8 e (E3)	897.3	700.63	511.7	407.5	315.64	194.27	120.4	0					
9 f (F3)	988.1	791.42	602.5	498.32	406.44	285.1	211.2	90.8	0				
10 fis (F#3)	1080.0	883.3	694.4	590.2	498.31	376.95	303.1	182.7	91.9	0			
11 g (G3)	1186.9	990.2	801.3	697.1	605.2	483.83	409.95	289.56	198.76	106.9	0		
12 gis (G#3)	1306.4	1107.8	918.9	814.7	722.8	601.42	527.54	407.15	316.35	224.48	117.6	0	
13 a (A3)	1389.26	1190.66	1001.75	897.56	805.68	684.31	610.43	490.04	399.24	307.37	200.48	82.9	0
14 b (Bb3)	1513.26	1314.66	1125.75	1021.56	929.68	808.31	734.43	614.04	523.24	431.37	324.48	206.9	124.0
15 h (B3)	1580.66	1382.1	1193.15	1088.96	997.08	875.71	801.83	681.44	590.64	498.77	389.88	274.3	191.4
16 c′ (C4)	1698.76	1500.2	1311.4	1207.06	1115.18	993.81	919.93	799.54	708.74	616.87	507.98	392.2	309.5
17 cis′ (C#4)	1783.76	1585.0	1396.2	1291.86	1199.98	1078.73	1004.73	884.34	793.54	701.67	592.78	477.0	394.3
18 d′ (D4)	1907.26	1708.5	1519.7	1415.36	1322.48	1202.23	1128.23	1007.84	917.04	825.17	716.28	600.5	517.8
19 es′ (Eb4)	1975.76	1777.0	1588.2	1483.86	1390.98	1270.73	1196.73	1076.34	985.54	893.67	784.78	669.0	586.3
20 e′ (E4)	2078.86	1880.1	1691.3	1586.96	1498.08	1373.83	1299.83	1179.44	1088.64	996.77	887.88	772.1	689.4
21 f′ (F4)	2197.06	1998.3	1809.5	1705.16	1612.28	1492.03	1418.03	1297.64	1206.84	1114.97	1006.08	890.3	807.6
22 fis′ (F#4)	2275.76	2077.0	1888.2	1783.86	1690.98	1570.73	1496.73	1376.34	1285.54	1193.67	1084.78	969.0	886.3
23 g′ (G4)	2398.86	2200.1	2011.3	1906.96	1814.08	1693.83	1619.83	1499.44	1408.64	1316.77	1207.88	1092.1	1009.4

(continued)

Table 4 (continued)

	1 G (G2)	2 A (A2)	3 H (B2)	4 c (C3)	5 cis (C#3)	6 d (D3)	7 es (Eb3)	8 e (E3)	9 f (F3)	10 fis (F#3)	11 g (G3)	12 gis (G#3)	13 a (A3)
24 gis′ (G#4)	2498.26	2299.5	2114.3	2006.36	1913.48	1793.23	1719.23	1598.84	1508.04	1416.17	1307.28	1191.5	1108.8
25 a′ (A4)	2597.86	2399.1	2213.9	2105.96	2013.08	1892.83	1818.83	1698.44	1607.64	1515.77	1406.88	1291.1	1208.4
26 b′ (Bb4)	2702.06	2503.3	2318.1	2210.16	2117.28	1997.03	1923.03	1802.64	1711.84	1619.97	1511.08	1395.3	1312.6

Table 5 Scheme of quarter-comma meantone temperament for 12 notes/keys per octave

–2	$f^{\#}$ ….	$c^{\#}$ ….	$g^{\#}$			
	|	|	|			
–1	d ….	a ….	e ….	b ….	$f^{\#}$...	$c^{\#}$
	|	|	|	|	|	|
0	b_b ….	f ….	c ….	g ….	d ….	a
				|	|	|
+1				e_b	b_b	f
	+11	+5.5	0	–5.5	–11 cents deviation from pure fifth	

between the pure major third 5/4 (386 cents) and the Pythagorean major third 81/64 (408 cents) is distributed to intervals of fifths and fourths whereby the fifths are narrowed and the fourths are widened, respectively. Tuning four pairs of just major thirds b^b – d –f#, f – a – c#, c – e – g#, e^b –g – b with c taken as the center, one obtains the following scheme in which just thirds are written in the vertical and connected by the sign | while the narrowed fifths are marked …. The signs 0, −1, −2, +1 denote comma differences of tones/pitches relative to tones in the 0-series of fifths.

If the system would be expanded beyond 12 pitches and tones per octave, continuation would be possible to the right of the bracket (but also on the left side). In ¼-comma meantone temperament that came into use in the 16th century (see [54]), the fifth above *c* is 5.5 cents too narrow while the fifth below *c* (equal to the fourth above *c*) is 5.5 cents wide. The error margin increases by c. 5.5 cents per fifth (a quarter of the 'syntonic comma' of 21.5 cents); 12 keys tuned to this scheme result in the chromatic scale shown above. Characteristic of this scale besides the just major third and a good approximation to the just minor third (ratio 6/5, 316 cents) as well as the just major sixth (5/3, 884 cents) is that it offers two distinct sizes for chromatic and diatonic semitones, whereby both come close to just intonation intervals in which the chromatic semitone (25/24) has only ca. 71 cents while its diatonic counterpart (16/15) is nearly 112 cents wide. Taken together, both semitones add up to a minor whole tone (25/24 * 16/15 = 10/9) while the major whole tone (ratio 9/8, derived from two fifths minus one octave, i.e. 3/2 * 3/2 * 1/2 = 9/8) spans 204 cents. The difference of course is the comma 81/80 = 21.5 cents. The meantone can be found both as the geometric mean between the two whole tone sizes (the exact value being 192.855 cents) and from a progression in narrowed fifths (see Table 5). In fact, the meantone *d* halves the just major third *c* – *e* and is one of the compromises one has to make to obtain as much as 8 just major thirds with only 12 notes and keys (on a keyboard) per octave.

Though the first semitone is missing in the Dumery carillon (as was often the case in historic instruments to cut costs for one expensive large bell that was rarely needed musically), one can see that several of the scale steps as expressed by their f_1 and respective cents approximate the ¼-comma meantone scale fairly well. There is some further evidence in this direction from the tuning of the prime partials

(see below). However, there are also some significant deviations from this scale type, making it difficult to decide whether or not meantone temperament was the model for the actual tuning of the carillon. In this respect, one has to take the overall spectral structure of the bell sounds into account. Because of the inherent inharmonicity, it seems problematic to represent each complex sound by f_1 alone (as could be done, at least in principle, with harmonic complexes where $f_n = nf_1$ and where $A_n \sim 1/n$, $n = 1, 2, 3, \ldots$).

It is worth noticing that bell founders and campanologists alike often consider the tuning of a set of bells in terms of their respective 'strike note'. The strike note (from all we know, see Sect. 4) is a subjective virtual pitch difficult to measure in an objective way. Therefore, bell tunings have been given by taking the nominal (the partial situated ideally a perfect octave above the prime) as a reference (hence the name 'nominal'). Similarly, the prime has served as a reference, assuming that the prime is exactly one octave below the nominal. These assumptions may be justified in practice since, in most minor third/octave bells manufactured and tuned with appropriate care, the interval between the nominal and the prime is very close to a perfect octave. However, objective measurements and data from psychoacoustic experiments show that the pitch attributed to the strike note is not always exactly an octave below the pitch of the nominal. Moreover, the pitch of the strike note must not coincide with that of the prime partial (for such data, see e.g. [40, 51, 55, 56]).

4 The Strike Note of Bells and Carillon Tuning

Subjects since long have experienced a certain component in the sound of a bell immediately at the onset that appeared different from partials such as the hum, tierce and nominal both in sound quality and duration. Whereas the low partials, when heard individually, have a rather soft sound quality and decay slowly (what holds true in particular for the hum), there is a component in the sound which most listeners usually describe as sharp and metallic in timbre, and short in duration. Because of these attributes, it was at times believed this component resulted from the clapper impact on the bell's wall, and hence from a metal ball striking a metal surface. Though the contact of clapper and bell may give rise to some transient noise, the contact time in fact is very short (about 1 ms, see above) while the component heard as the so-called strike note seems to last, in most instances, for a fraction of a second so as to yield a more or less clear sensation of a pitch. As can be inferred from psychoacoustic data, a stable sensation of pitch from complex sounds emerges within tens of milliseconds (according to an estimate in de Cheveigné [57, 205], an average value of 66 ms seems reasonable). If the strike note is perceived as an identifiable pitch, this implies relevant sensory information derived from auditory processing of sounds must be present for some time, which can be estimated as covering probably $50 \leq t \leq 200$ ms. A time span of this size of course leaves the bell/clapper contact noise unlikely as a source for the strike note pitch.

4.1 The Strike-Note as a Virtual Pitch

Since the pitch of the strike note subjects perceived could not be attributed to a single spectral partial, the strike note was labelled 'imaginary' to mark the difference from acoustically 'real' partials [58]. Several hypotheses on the nature of the strike note were issued in studies published before ca. 1950 (summarized in [11, 51, 55]), some of which addressed the strike note as a difference tone resulting from a combination of strong partials and as a product of aural distortion. It became clear that the strike note is not a physical component somehow contained in a complex inharmonic sound but a perceptual phenomenon resulting from auditory and possibly neural processing of such complex sounds. In regard to pitch perception, a pitch resulting from any strong spectral component can be labelled 'spectral pitch' while a pitch resulting from a combination of spectral components can been called 'virtual' if it gives rise to a pitch subjects locate at a frequency where no significant spectral energy is found (see [59]). For example, in a harmonic complex such as generated from bowing a string, one can produce one spectral pitch when the fundamental f_1 is by far the strongest partial in a spectrum where partial amplitudes roll off at, e.g., $A = 1/n$, n = harmonic partial no. 1, 2, 3, …. Typically, subjects will hear another spectral pitch one octave above f_1 when the second partial (f_2) is as strong or even stronger in amplitude than f_1. A virtual pitch, on the other hand, results from a combination of several harmonic as well as (depending on conditions) inharmonic spectral components. Subjects in many instances locate a virtual pitch at a frequency where no spectral energy indicative of a partial is found; however, coincidence of a spectral and a virtual pitch is possible (a case relevant for bell sounds, see below).

A reasonable explanation for the strike note understood as a virtual pitch was issued by Schouten [60], also Schouten and t'Hart [61] who argued that several periodicities from a series of harmonic partials combine into a common period, which is enough to produce a pitch percept equivalent to the frequency with which this common period repeats per second. The repetition frequency f_0 of a complex waveshape resulting from the superposition of several consecutive harmonics equals the fundamental frequency f_1 of that harmonic series and gives rise to a pitch corresponding to f_1 even though f_1 might be missing in the signal altogether. For example, harmonic partials 3, 4 and 5 taken with equal amplitudes in either sine or cosine phase combine into their common period $T_0 = 1/f_0$ which equals the period (in ms) corresponding to the fundamental frequency f_1 of the harmonic series chosen. Hence the repetition frequency f_0 derived from the common period is equal to f_1. Since $f_0 = f_1$, the repetition frequency f_0 is substituted for f_1, which is the 'missing fundamental'. The relation sketched here holds even if the harmonics are not linked in a series like, for example, harmonics 3, 5 and 9 which (as harmonic partials in sine phase) combine into a periodic waveshape whose repetition frequency f_0 of course equals f_1 (see Fig. 11).

If musically trained subjects are presented with a complex sound composed from superposition of several (consecutive or non-adjacent) harmonics, they can be

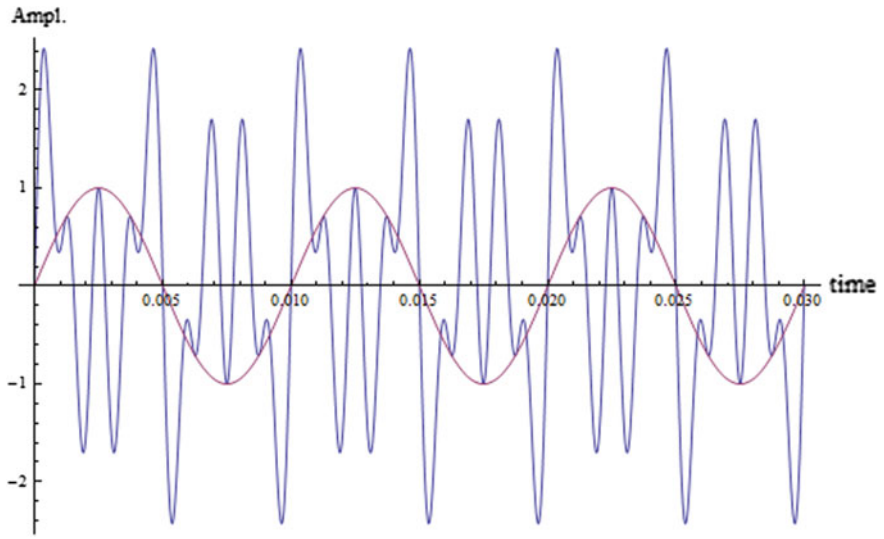


Fig. 11 Periodic waveshape (*blue*) resulting from superposition of harmonics 3, 5 and 9 (300, 500, 900 Hz). Repetition frequency f_0 = 100 Hz (*red*) of the complex waveshape equals f_1 of the harmonic series

expected to sing or hum a pitch they believe represents that complex. Let, for example, the complex comprise three partials of equal amplitude at 700, 900, and 1100 Hz, respectively. If this complex is played back to a subject at ca. 70 dB via loudspeakers for ca. 2 s, and the subject is asked to sing or hum the pitch he or she perceives immediately after the stimulus ends, it is likely that she or he will produce a sound that itself has a f_1 and/or a f_0 at 100 Hz as shown in Fig. 12. To be sure, the stimulus (spectrogram from three harmonic partials) and the response (pitch track from AC analysis) here have been plotted into one graph while the stimulus and the response in fact are two different sounds not overlapping in time.

It is known from experiments that the salience of the pitch corresponding to f_0 of a harmonic complex depends on the harmonic number of the partials as well as on the frequency region in which these partials fall (see [62], [63], Chap. 7). In general, salience of a virtual pitch based on f_0 is greater for complexes comprising low harmonic partials which can be 'resolved' by the cochlear filter bank (preferably, consecutive harmonics such as 3, 4, 5 or 4, 5, 6) and for f_0 falling into a frequency band ranging, roughly, from about 100–500 Hz (which means partial frequencies ranging from about 300 to 3000 Hz). However, it was found that also groups of higher harmonic partials which cannot be resolved aurally into their constituents can give rise to virtual pitches (cf. [64]). Though such groups of higher harmonics establish a common periodicity, the pitch salience for f_0 from such stimuli is significantly lower in comparison to the salience of low spectral pitches and f_0 pitches resulting from low harmonics.

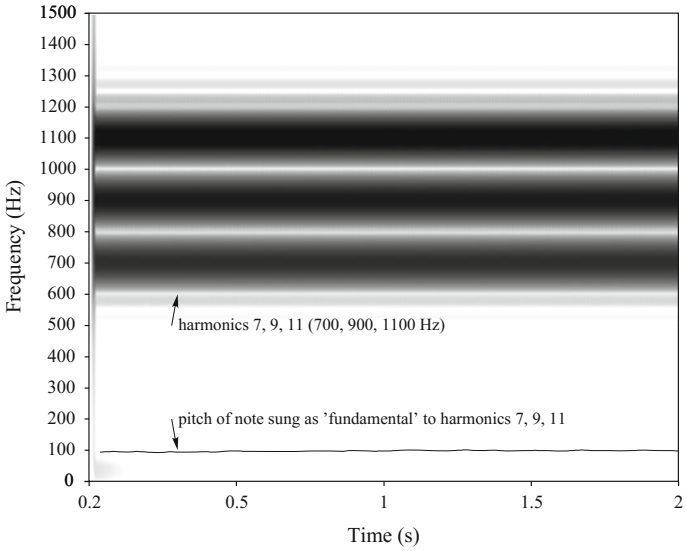


Fig. 12 Spectrogram of a harmonic complex {700, 900, 1100 Hz} and pitch track at about 100 Hz derived from the tone sung by a subject after listening to the complex

A point that needs some comment is the nature of pitch perception based on either f_1 or f_0. In principle, spectral pitches corresponding to pure (sine or cosine) tones of a given frequency result from direct sensory information, namely stimulation of the basilar membrane (BM) at a certain place. The relation of frequency to BM place is known as tonotopic, resulting in a cochleatopic map. Though the area of BM excitation is not indefinitely small even at low sound levels (and broadens significantly with SPL > 40 dB, see [65], there are mechanisms suited to transform the place information into an unambiguous neural signal that preserves the frequency of the stimulus as a basic correlate of 'tone height' (which in turn is a basic constituent of pitch; see [59], Chaps. 9–11, [66]). Since $T = 1/f$, the period corresponding to a certain stimulus frequency is also contained in the neural spike train. In regard to pitch salience, for complex harmonic sounds which include the f_1 partial, and where f_1 often is strong in amplitude, essential sensory information comes from a spatial pattern of cochlear excitation. In addition to place information, a harmonic complex comprising a number of partials including f_1 provides temporal information since the virtual f_0 'fundamental' corresponding to the period of the complex waveshape not only equals but also reinforces the f_1 pitch.

In contrast, for harmonic complexes lacking f_1 (and maybe also other low partials), much of the information attributable to spatial excitation on the BM is missing. Hence, the f_0 pitch must be inferred from temporal information pertaining to the overall periodicity of the stimulus. In certain respects, different mechanisms of excitation and processing which are behind f_1 and f_0 pitch perception can account for differences in pitch salience. While a strong f_1 component in a complex harmonic sound in general gives rise to an immediate sensation of this partial and a

clear pitch perception, many harmonic complexes with 'missing fundamental' and concentration of spectral energy towards higher partials (i.e., the centroid moves up) in fact not only lack that particular sensation but also convey a somewhat weaker pitch. In regard to perception (which includes cognitive assessment of what has been perceived), a virtual f_0 'fundamental' appears as being 'implied' by the sensory information derived from peripheral processing of a stimulus lacking f_1 while the f_1 fundamental in a homogeneous harmonic complex (i.e., the series of partials is complete with $A_n \sim 1/n$) seems to result from an immediate and 'automatic' response.

Schouten's concept of 'temporal' pitch (see [60, 62]) based on the common period constituted from several or even many harmonics explains perception of the so-called missing fundamental and served as a hypothesis for the strike note in bell sounds. Schouten [60], also Schouten and 't Hart [61] suggested that partials nos. 5, 7 and 10 of a carillon bell would have frequency ratios close to 2:3:4, and that the strike note close in frequency to the prime (partial no. 2) would thus be perceived as a kind of missing fundamental. In Schouten's scheme (see also Table 2), the relevant partials are the nominal (or octave above the prime), the twelfth, and the double octave, which usually are strong spectral components. With these partials present in the spectrum, and provided they carry sufficient energy, perception of a strike note at about the frequency location of the prime seems feasible. In fact, in a number of experiments the strike note was found to result from a combination of nearly harmonic partials contained in the sound radiated from bells (see [55, 51, 56]). The location of the strike note often (but by no means always) was found close to the prime. For a number of bell sounds, the strike note was either below or (more frequently) above the prime. Also, in the sounds projected from some bells more than one strike note could be identified (a second strike note often appears a fourth or major third higher than the first). The frequency position of a strike note can be shifted by either manipulating the frequencies of relevant components in synthesized bell sounds [55] or in the sound of a real bell [67]. In these experiments, it was found that the octave (nominal), the twelfth and the upper octave partials are the most important contributors to strike note pitch.

Taking the 'strike note' as equivalent to perception of a 'missing fundamental' (see [61]), also a frequency location close to the hum would be possible. To illustrate the case, one may design an 'ideal bell' with partial frequencies like 100, 200, 240, 300, 400, 500, 600, 800, 1200, 1600 Hz and amplitudes similar to those found in the sound of a real bell.

For such a sound synthesized from Fourier components (for simplicity, all in cosine phase, see Fig. 13), the most likely strike note pitch frequency derived from temporal and spectral information would be 100 Hz as was determined with two pitch detection algorithms based on autocorrelation (AC, [68, 69]) and another pitch algorithm based on subharmonic summation (SHS, [70]). Thus, the most likely strike note pitch in this case coincides with the frequency of the lowest spectral component, f_1, which is found as the common denominator fitting best to the partly incomplete and partly not quite harmonic series of partial frequencies listed above. If, however, the sound of a real bell is analyzed which may contain partials with

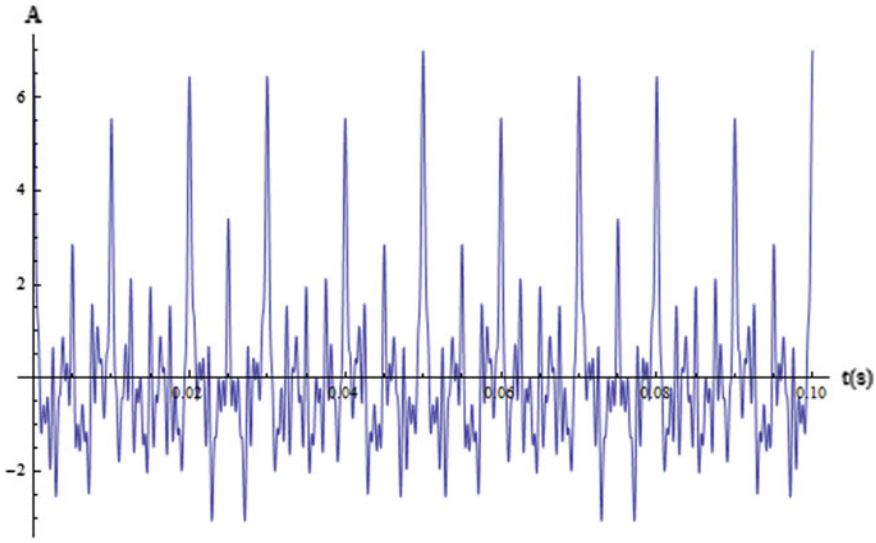


Fig. 13 Synthesized minor-third bell sound, oscillogram for 100 ms. A periodicity at $T = 10$ ms ($f_0 = 100$ Hz) is still clearly visible

frequencies in the same range as those used for the synthesis plus a significant number of more or less inharmonic components interspersed between the 'principal partials' as well as in higher octaves (see Fig. 8), the result may be different. For example, spectral analysis of the sound of bell no. 3 of the Dumery carillon shows that $f_1 = 122.06$ Hz. The pitch detection algorithm based on AC yields ca. 124 Hz measured at both $t = 0.5$ and $t = 1$ s from onset. Thus, the autocorrelation method taken to determine the pitch of the strike note in this case as well as for most of the 26 Dumery bells calculates a frequency close to the fundamental (and in several cases also a frequency close to half of f_1). On the other hand, the subharmonic summation algorithm (SHS) for the same sound finds 246.71 Hz at $t = 0.5''$, 245.5 Hz at $t = 1''$, and 245.31 Hz at $t = 2''$. Hence, the pitch frequency that might represent the strike note in this case is close to, yet not identical with, the frequency of the prime (which in bell no. 3 is a doublet comprising two components at $f_{2a} = 243.75$ Hz, $A_{2a} = 66$ dB and $f_{2b} = 244.01$ Hz, $A_{2b} = 72.1$ dB). Since campanologists tend to locate the strike note at, or close to, the frequency and spectral pitch of the prime, we may accept the result of the SHS measurement as correctly representing the frequency of the strike note pitch. Such a conclusion seems plausible, on the one hand, though it probably simplifies matters, on the other.

4.2 Strike Note, Pitch and Timbre

To understand the issue of strike note pitch as distinct from spectral pitches more closely, one has to take the timbre of sounds and the interdependence of pitch and timbre into account. In much of the relevant literature, the term 'timbre' refers to the spectral composition of a sound including its temporal and dynamic modifications while 'sound colour' rather relates to the shape of the spectral envelope and the position of the spectral centroid (for a detailed discussion of both concepts, see [66]). Hence, the sound of a bell can be described in terms of 'timbre' for the change it shows in spectral energy distribution over time while a sine tone or a harmonic complex in the steady-state can be described in terms of a 'sound colour'. For a sine tone of given frequency and amplitude, there is in general one unambiguous sensation of pitch depending on the stimulus frequency. Note that low to medium sound levels (30–60 dB) exert almost no influence on the pitch (see [71], Chap. 5). There is another attribute of sensation known as tonal brightness which, for pure tones, again depends on stimulus frequency. The 'colour' of simple stimuli such as sine tones does not change significantly within one octave (say, 200–400 Hz) where a smooth increase of relative brightness of the sound with frequency will be observed (the level being held constant). A similar effect can be expected for harmonic complex sounds where all components are coupled in phase and where the amplitudes of partials conform to $A_n = 1/n$, n = harmonic number or roll off at a similar rate per octave (see above). For such sounds, pitch is conveyed unambiguously by both f_1 and f_0 information, and their 'colour' does not change significantly if all partials of a harmonic complex are shifted by a musical interval (say, a major third or even a fifth) up or down while their relative amplitudes are maintained. In effect, this implies a shift of the spectral envelope which causes changes in the relative brightness sensed since the spectral centroid (see [72]) moves up or down in accordance with the shift.

In classical hearing theory [73], the fundamental of a harmonic spectrum was regarded as the main carrier of pitch information (sensation of f_1 at a certain place on the BM), and the energy distribution from the remaining partials as largely determining the sound colour (also labelled 'tone colour'). Hence, pitch as a function of f_1 and sound colour conceived mainly as a function of the shape of the spectral envelope become separable as percepts even if some interaction is taken into account. Of course, the f_1 partial not only defines 'fundamental' spectral pitch but also contributes energy to whatever sound colour. Furthermore, concentration of spectral energy above the f_1 fundamental can provoke a pitch shift (for harmonic spectra, a shift by an octave is a likely case). The Helmholtz model of pitch and sound colour sensation obviously relates to the harmonic line spectrum such as sketched in Fig. 14. For most string and wind instruments (chordophones, aerophones), sounds usually exhibit a strong f_1 partial suited to determine pitch. Amplitudes of higher partials often roll off at a significant rate (e.g., −6 dB/octave) relative to the f_1 amplitude. The number and intensity of the partials above f_1 then would determine the 'colour' of a particular sound.

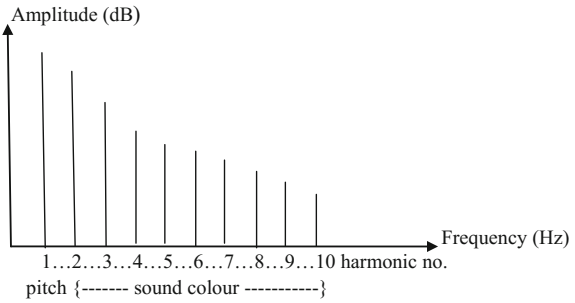


Fig. 14 Harmonic line spectrum as a model of pitch and sound colour sensation

In bell sounds, temporal and spectral structure, in general, are much more complex depending on the regime of vibration (see above). Complexity of sound structure in bells can show strong effects on the sensation and perception of pitch and timbre. The decisive factor for the dynamics and the timbre of carillon bells of course is the magnitude of the impact force with which the clapper excites the bell. In carillons offering a pedal-board (as in Bruges), the player can accelerate the clapper rapidly when kicking one of the pedals, thereby applying a strong impact force to the respective bell. Such a playing technique, which is quite common for bourdon notes that are intended to be perceived as marking the beginning of a phrase, has the effect of eliciting very many eigenmodes in a bell; from the vibration a complex pattern of more or less harmonic partials combined with inharmonic components (not to forget the split of partials into doublets, see above) results in the spectrum. Taking the correspondence between the periodicity of a signal in the time domain and its harmonicity in the frequency domain into account (a correspondence explained by the Wiener-Khintchine theorem; see e.g. [74]), one can measure the degree of periodicity in a signal in the time domain by computing the harmonic-to-noise ratio (HNR, see [68]). The concept rests on viewing a signal as consisting of a number of harmonic partials which correlate among each other while there may be other components (such as noise) uncorrelated with the harmonics. The algorithm treats both parts of the signal in calculating the HNR expressed in dB over time. A strictly periodic signal with no noise can yield high dB ratings (e.g., a sawtooth wave generated with Mathematica® and analyzed with this HNR algorithm reaches some 60–65 dB) while in particular the onset of a bell sound yields quite low dB readings due to spectral inharmonicity even if only a medium impact force is applied to the bell. Later in the sound, periodicity of the signal increases since many of the transient inharmonic partials decay rapidly and then may vanish completely. Consequently, the HNR yields higher dB after about one second of sound has elapsed. To illustrate the case, Fig. 15 shows the HNR (dB) of bells no. 9 and no. 6 from the Dumery carillon.

The rather inharmonic sound structure of bells at the onset and within the first 0.5 s is of consequence also to sensation and perception. First of all, sounds recorded in the vicinity of a bell can reach a fairly high SPL immediately after the clapper impulse is transmitted to the bell and has excited a large number of

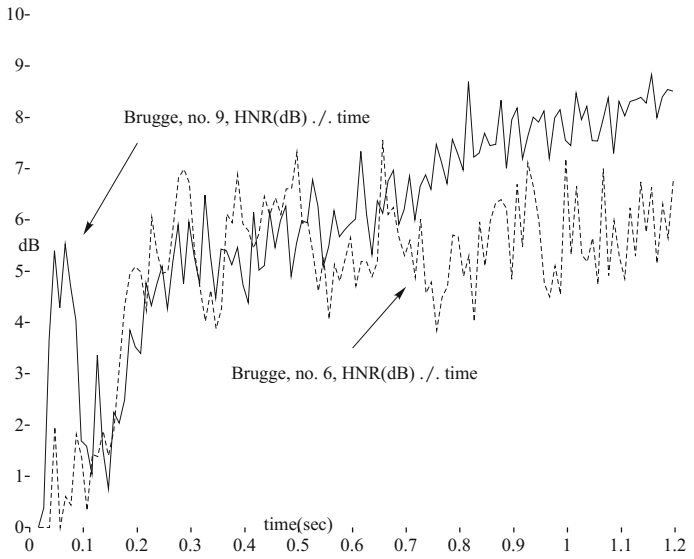


Fig. 15 HNR (dB) over time for sounds from bell no. 9 (*solid line*) and bell no. 6 (*dashed*)

eigenmodes. Second, the broad spectrum covered in the sound by many nearly harmonic as well as inharmonic partials in turn produces broadband excitation on the basilar membrane of the cochlea. The effect of both parameters is evident in Fig. 16 which combines the cochleagram for a sound recorded from bell no. 9 with the plot of the corresponding sound intensity over time (which peaks close to 80 dB).

Sound intensity and wide spectral energy distribution including inharmonic components shortly after onset account for the sensation of a relatively loud sound that to campanologists and musical listeners alike appears metallic and sharp in timbre (see [41, 51, 55]). Sensation of such a timbre may affect pitch perception to some degree since both cannot be neatly separated. In fact, pitch and timbre closely interact in particular in sounds which have inharmonic spectra such as recorded from Javanese and Balinese metallophones or gong chimes (see [32, 75]) where quite often the spectral component corresponding to the lowest mode of vibration is not the strongest in radiated sound level. In addition, sensation of pitch can be blurred from groups of inharmonic components, which are too close in frequency to be 'resolved' on the BM level. Note that such groups of inharmonic components interact so as to give rise to AM and a sensation of roughness and beats. Furthermore, with increasing degree of inharmonicity of the spectrum overall periodicity of the time signal decreases, to the effect that sensation of f_0 pitches is hampered (see [76]). Taking sounds such as recorded from Indonesian gong chimes (e.g., the *bonang* of Java or the *trompong* of Bali), many if not most Western listeners perceive a clangy, metallic sound rather than a musical tone distinct in

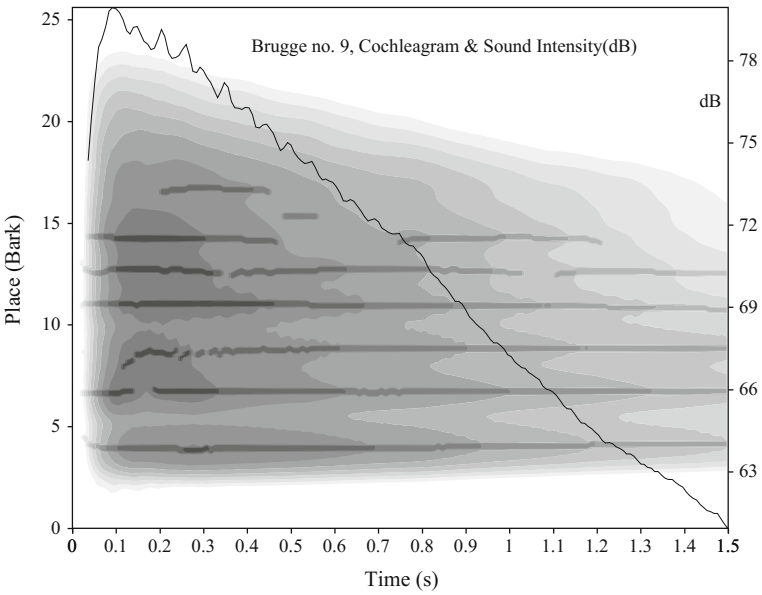


Fig. 16 Cochleagram and sound intensity (dB) as a function of time, bell no. 9

pitch. In the experience of many musically trained subjects (but unfamiliar with these sounds) no clear distinction between pitch and timbre seems possible.

Though many western swinging and carillon bells have spectra which include partials that can be assigned to (one or several) harmonic series, there are still enough inharmonic components to cause ambiguity of pitch perception (see [40, 56]). In general, there seems to be a clear correspondence between spectral inharmonicity and ambiguity of pitches subjects perceive from the sounds of bells. The ambiguity apparently results from a concurrence of several pitches, both spectral and virtual, and also from the interaction of pitch and timbre that seems to affect perceptual analysis of the strike note. While the pattern of partials involved in good minor third/octave bells in general facilitates perceiving a 'main pitch' (which in many cases is the strike note close to the prime), the sharp metallic timbre attributed to the strike note may perhaps diminish the salience of the strike note pitch.

Among the spectral partials of swinging and carillon bells, the minor third because of its level in the spectrum and because of the interval it forms with the prime and the nominal almost always can be identified by listeners as a spectral pitch. In addition, listeners often can hear (and reproduce by humming or singing) a few other partials such as the nominal if bell sounds are presented in isolation. Taken together, the prime and the strike note also form a possible pitch area. Hence, several strong spectral partials plus the strike note may give rise to a sensation of multiple pitches in each complex bell sound. As a hypothesis, one may expect subjects with some musical background to be able to distinguish several spectral

Table 6 Number of pitches perceived per bell, 15 bells, n = 81 subjects

No.	Median	Mean	SD
1	2	2.63	1.07
2	2	2.73	1.09
3	3	2.65	1.07
4	2	2.3	0.86
5	2	2.43	1.14
6	2	2.46	1.17
7	2	2.39	1.04
8	2	2.53	1
9	2	2.44	0.9
10	2	2.28	0.78
11	2	2.2	0.95
12	3	2.99	1.16
13	2	2.48	1.08
14	2	2.23	0.93
15	3	2.65	1.06

partials as well as the strike note that are perceived as separable pitches. To test this hypothesis, we presented the sounds from the first 15 (no. 1–15) of the 26 Dumery bells to a total of 81 subjects, most of them students in musicology at Hamburg in their first year. All sounds were digitally normalized to a level of -6 dB and were played back from hard disc in a class room suited to musical performance and recording via a stereo system comprising high quality loudspeakers at an SPL of ca. 75–80 dB(C) measured 1 m from the source. The level was chosen to offer conditions as one would experience if standing not too far from to the actual bells. Each sound (average duration ca. 4 s) was offered twice, with a break of 1 s in between. Subjects were asked to note the number of pitches perceived per bell sound in a questionnaire. The data for 15 bells (nos. 1–15) are given in Table 6.

Calculated over 15 bell sounds and 81 subjects, the mean is 2.49 and the SD is 1.04. Hence, the average number of pitches subjects perceived from single bell sounds recorded from the historic Dumery carillon in this experiment was from two to three.

To explore the ambiguity of bell sounds in regard to pitch perception further, we conducted an interval identification test with the same 81 subjects. For this task, 20 intervals were formed from the sounds of the bells nos. 1–13 as listed in Table 7. There were 10 intervals in upward direction and 10 intervals in downward direction. The two sounds, A and B, forming an interval were played in succession (A → B) where each sound lasted for ca. 3–3.5 s. The gap (silence) between two sounds was 0.5 s. Each pair of sounds representing a certain interval was repeated once after a short break of 0.5 s, thus the sequence was sound A: silence: sound B: silence: sound A: silence: sound B. Subjects were asked to state the size of each interval either with a musical term (e.g., ‘major third’) or by expressing the interval by the number of semitones it spans (e.g. 8 = minor sixth). In this regard, the exact

Table 7 20 intervals presented with bell sounds as stimuli; no. of hits

Trial no.	Bells no.	Musical interval	Direction	No. hits
1	1 → 4	Fourth	Up	27
2	8 → 6	Major second	Down	19
3	2 → 10	Major sixth	Up	15
4	11 → 4	Fifth	Down	32
5	2 → 5	Major third	Up	26
6	13 → 12	Minor second	Down	50
7	3 → 6	Minor third	Up	41
8	11 → 5	Tritone	Down	5
9	4 → 12	Minor sixth	Up	19
10	13 → 8	Fourth	Down	27
11	7 → 4	Minor third	Down	30
12	1 → 10	Major seventh	Up	19
13	3 → 5	Major second	Up	50
14	12 → 2	Major seventh	Down	18
15	7 → 13	Tritone	Up	14
16	6 → 10	Major third	Up	29
17	2 → 8	Fifth	Up	28
18	3 → 4	Minor second	Up	30
19	13 → 5	Major sixth	Down	11
20	11 → 3	Minor sixth	Down	18

size of the interval relative to a peculiar tuning was not considered (in Pythagorean tuning, the minor sixth has 792.2 cents, in ET12 it has 800 cents, and in just intonation the interval size is 813.7 cents). Hence, the decisions to be made by subjects were between basic musical intervals such as minor second, major second, minor third, major third, etc. Even though subjects were not explicitly asked to mark also the direction of the interval (up, down), most of the subjects did include such information (e.g., by noting "7↑" or "11↓"). The 20 intervals presented as stimuli are listed in Table 7.

Of the responses collected from 81 subjects, 76 lists containing their respective interval ratings were usable. The number of correctly identified intervals ('hits') per trial is stated in the last column of Table 7. The sum of correctly identified intervals is 508. Since 76 subjects had to make 76 × 20 = 1520 interval judgements, the number of 508 hits means a fraction of only 33.4 % of the judgements was correct. On average, subjects identified 6–7 intervals out of 20 correctly (median = 6, mean = 6.75, SD = 4.1). The relatively small number of 'hits' may reflect a general difficulty for subjects to judge musical intervals played with sounds that for most of the listeners were unfamiliar and, at least for some of the intervals, highly ambiguous if not contradictory (in particular, the major third played with two minor third bells). Inspection of the data reveals that the 76 subjects differed significantly

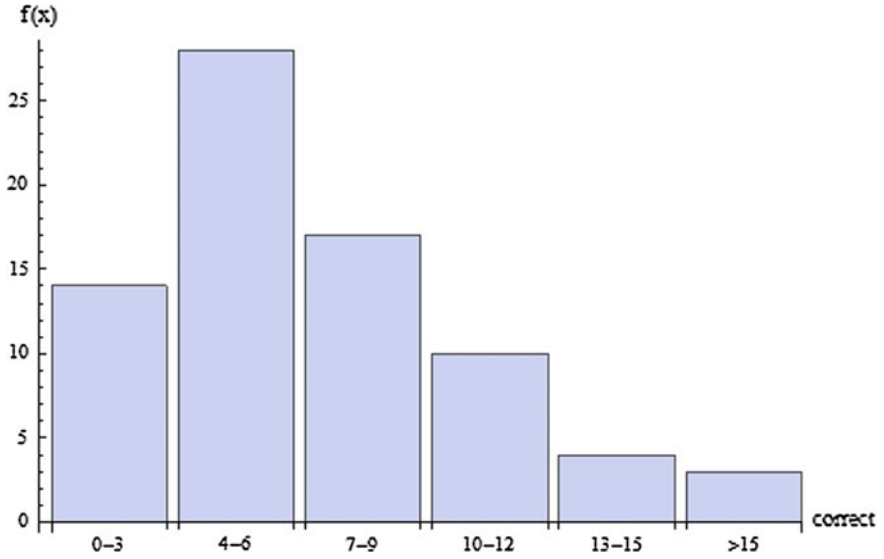


Fig. 17 Frequencies of correctly identified intervals (81 subjects) ordered into 6 classes

in their performance since some achieved a high rate of correct judgements (the maximum was 17 correct out of 20 intervals) while others evidently had great difficulties to identify musical intervals when presented with minor-third bells. However, there was a fairly large number of 'near misses' (judgements falling ± 1 semitone off the correct interval). Further, a number of judgements seemingly reflected confusion errors known also from experiments with so-called 'possessors of absolute pitch'. One such confusion error is that even musically trained subjects at times mistake a pure fifth for a perfect fourth (et vice versa); also, subjects may take a major sixth for a major third, or a minor sixth for a minor third by judging the relative consonance in both pairs of intervals. Ordering the results from 76 subjects into six classes according to the number of hits, the bar chart shown in Fig. 17 can be plotted.

Results from our interval identification task, albeit exploratory in nature, once again underpin pitch ambiguity in carillon bell sounds. The partly inharmonic structure of each minor-third bell sound not only hampers pitch perception but, as a consequence, also interval identification. Since subjects in general perceive more than one pitch from each carillon bell, the difficulty lies in assigning sounds that give rise to several pitches to the steps of a musical scale that in itself is one-dimensional in several respects.

First of all, 'musical pitch' basically is defined by log frequency as note names like $A_3 = 220$, $C_3 = 261.5$ or $A_4 = 440$ Hz imply. Hence, a scale of musical tones in ET12 for one octave (A_3 to A_4) can be plotted as a function of $\log(f)$ like Fig. 18.

In this scale, the frequencies can either represent sine tones or the f_1 of harmonic complex tones. Conventional (staff) notation represents musical tones

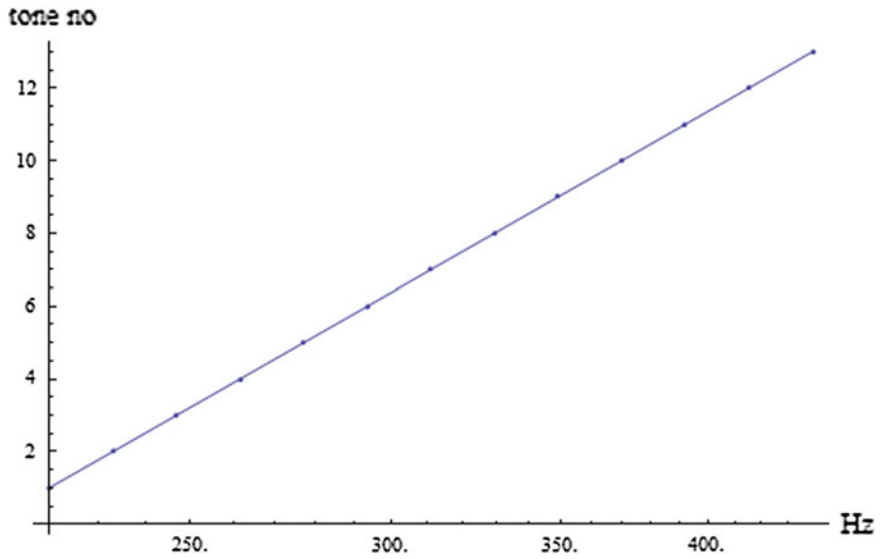


Fig. 18 Musical scale (A_3 to A_4) defined by 13 frequency values (*dots*)

(i.e., harmonic complex tones such as produced by singers or by wind and string instruments) by their respective f_1 while it disregards other spectral partials (as such are believed to be relevant not for pitch but for 'sound colour', see Fig. 14).

Second, the musical scale formed from a number of tones per octave is also conceived as one-dimensional since, in conventional music theory and music psychology, the scale steps are regarded as a category scale which may comprise k pitch categories (depending on musical culture, one finds different values for k). In regard to pitch production and perception, each category can be defined by a center frequency (as are marked by dots in Fig. 18) as well as by boundaries within which sounds with different f_1 shall be taken to represent the 'same' category (a 'pitch category' in certain respects corresponds to a Thurstonian scaling model with mean and variance, respectively; see [66]). Such a scale model works fairly well as long as one is dealing with sounds compliant with the harmonic line spectrum (Fig. 14), meaning (a) the f_1 partial of every harmonic sound (e.g., the musical 'tone' as produced on a chordophone or aerophone) is prominent and suited to determine f_1 pitch, and (b) is close to the (center) frequency defining the respective note and pitch category. Within the limits of each category, a direct relation between f_1 of a harmonic complex sound such as a tone from a wind or string instrument and a certain scale step can be established. For example, if a musical work contains a certain note (say, a G_4) a violinist or singer is expected to produce a harmonic complex tone with f_1 very close to 392 Hz (in ET12). However, a moderately detuned A_4 on a violin with f_1 at 434 Hz is still accepted, by most listeners, as 'representing' the note and the pitch category A_4 though f_1 here is flat by ca. 24 cents (a noticeable deviation yet within the limits of the pitch category).

For many inharmonic sounds, such a simple relation between f_1 component in the spectrum, perceived pitch and musical scale step is not at hand. The task for subjects dealing with sounds from bells (or, even worse, Javanese gong chimes) rather is to make a perceptual evaluation of where a sound that gives rise to several pitches may fit into a one-dimensional musical category scale. If several such inharmonic complex sounds are presented simultaneously and/or in a sequence, subjects have to perform many perceptual analyses and must make decisions as to the presumed pitch structure these sounds might 'represent'. In effect, the multiplicity of pitches induced from several inharmonic sounds increases the perceptual and cognitive workload in particular in a music listening situation where processing needs to be done almost in 'real time'. For actual performances of music on carillons, there is another factor that must be taken into account in regard to perception. Since every bell radiates sound for several seconds after the strike before significant damping takes place (Fig. 6), there is a temporal as well as a spectral overlap of complex inharmonic sounds. Therefore, with two or even three bells played simultaneously, pitch perception and recognition of melodic and harmonic textures can be quite difficult even if the music may be well known from other contexts (see [77]). Though listeners with a musical background can often identify certain songs rendered on a traditional carillon by perceiving their distinctive melodic and rhythmic features, the basic ambiguity of pitch (and also timbre) resulting from the partly inharmonic spectrum as well as from concurrent spectral and virtual pitches remains. Among the truly amazing experiences one may have with carillon music is listening to a piece full with major chords played with a carillon of minor-third bells. Many listeners may experience a certain incongruity or even discrepancy between the sound as sensed and the musical structure perceived as intended. In order to overcome this discrepancy, bells with a major-third spectrum have been designed for carillons (see [49]; swinging bells with a harmonic spectrum have been founded much earlier, see [55]). However, a certain degree of ambiguity may be experienced as appealing to listeners who might esteem a carillon as a musical instrument with a peculiar sound structure. Of course, there is a corpus of (in particular, traditional) music that fits to this sound structure. Various sources of the 17th and 18th centuries, respectively, indicate that music played on Flemish and Dutch carillons included many folk songs and hymns which often were elaborated in a characteristic two-part setting where the melody in the discant was played with ornamentation while the bass line consisted of longer notes (see [78]), thus taking the slow decay of sound level experienced in larger bells into account. In the 19th and 20th centuries, respectively, more of the virtuoso style known from piano and also organ playing was adopted by composers of carillon music as well as by carillonists. Playing music faster on a carillon and producing much more complex sonorities means, however, that the resulting sound mixtures can be so dense with harmonic and inharmonic components that perceptual analysis is very difficult, if possible at all.

Finally, a few tests were run to see which real or virtual component might be dominant in the pitch sensation and perception of musically trained subjects. The background to this issue is that several studies suggest that the acoustically real

prime and the virtual strike note are decisive in regard to pitch sensation and perception. The 'nominal' (as the name implies) has also been viewed as the partial that channels pitch perception. It should be noted that, in English campanology, the nominal in fact is the octave above the prime while recently German campanologists have labelled the strike note of a bell as the 'nominal' since the 'nominal pitch' (defining the position of a bell in musical scale) would be that of the strike note [79]. The assumption is that the frequency position of the prime partial is also that of the strike note, whereby the two components together constitute a dominant pitch percept defining the 'nominal' pitch of each bell. However, though this may be the case with many bells, one still should test whether pitch perception is as uniform as suggested.

There are several well-known methods used to test which pitch or which pitches subjects perceive when listening to the sound of bells (or to other sounds which are of interest). One is to make subjects match a sine tone to a test sound so that the two appear equal in pitch (see [40, 56, 59]). Sine tone adjustments can be repeated in case subjects perceive several pitches (one of which may be more prominent than the others and, consequently, be taken as the 'main pitch' evoked from a complex harmonic or inharmonic sound). Another method is to let subjects sing or hum the pitch they believe to have perceived from a test sound presented just before. Again, if several pitches are perceived, subjects may sing or hum tones so as to express these pitches. A third method is to make subjects match a harmonic or inharmonic complex to that of an harmonic or inharmonic test sound so that both appear 'equal' in pitch.

Each method has some advantage as well as possible drawbacks (for a comparison, see [51]). For example, a good reason for matching a sine tone to an inharmonic complex is that the sine tone itself is well-defined in pitch and easy to handle in measurement. However, there is a marked difference between a sine tone of constant 'sound colour' (which appears soft and smooth) and an inharmonic complex which may appear quite rough from AM and harsh in timbre. In particular, the difference between the strike note of bells, judged as metallic in timbre and short in duration, seems significant relative to the steady and soft sine tone. The method of singing what subjects regard as either the main pitch perceived from a bell or at least one of the separable and identifiable pitches is less affected by differences in timbre and has also the advantage that such a response can be uttered immediately after hearing a test sound (that is, by making use of echoic and short-term memory). This method, though, calls for subjects capable of singing a note at a distinct pitch, which in turn may require some musical training. The third method eliminates any significant difference in timbre but introduces unknowns into the experiment since a harmonic or inharmonic complex used for comparison with a test sound may itself give rise to several (virtual as well as spectral) pitches. If in an experiment subjects have to adjust an inharmonic complex (derived from Fourier synthesis as in Fig. 13) to an inharmonic test sound (say, the sound of a carillon bell or a Javanese *bonang* gong) which is likely to produce a multiplicity of pitch sensations, the synthesized complex can also be expected to give rise to more than one (spectral and/or virtual) pitch. Hence, synthesized inharmonic complexes need to be studied

with respect to their potential spectral and virtual pitches in beforehand of actual comparison to test sounds.

Taking the pros and cons of the three methods, singing a pitch or several pitches one perceives from listening to a test sound perhaps is a fairly reliable way to explore the issue. In a simple experiment, we asked 8 musically trained subjects (7 male, 1 female) to sing the pitches they perceived from listening to various bell sounds (Dumery bells nos. 1–12). The sounds were presented at low to medium level (ca. 40–50 dB) from a CD system and the responses were digitally recorded on hard disc. Spectral analysis and AC pitch tracking reveal that most responses of the male subjects aimed at producing a sound at a pitch that equals the frequency of the hum partial (rather than the prime or even the nominal). For example, the following correspondences were observed:

Bell no.	f_1 partial (Hz) bell	f_1 of sound sung as equivalent in pitch (Hz)
1	97.5	98.45
3	122	125.1
5	136.7	136.7
7	153	155.8
9	172.85	171.5
10	182.3	184
11	193.9	197

To be sure, f_1 of the vocal responses are averaged over time while the actual utterances showed some fluctuation in pitch (see Fig. 19). Among the responses there was one where a subject did sing a subharmonic (below f_1 of a bell) and very few responses showed subjects aiming at the prime (or, possibly, the strike note); one such response was recorded for bell no. 12 (see Fig. 19) from the female taking part in the experiment.

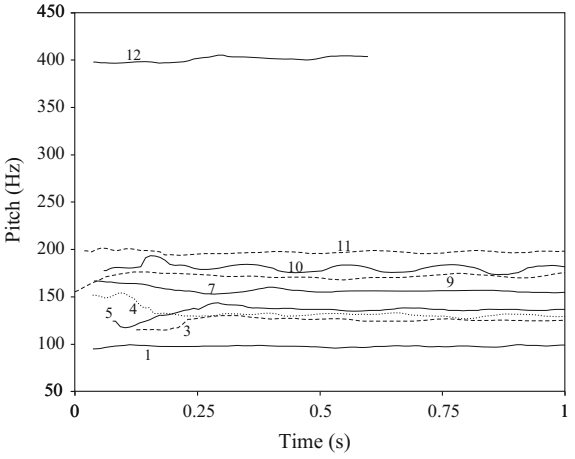


Fig. 19 Pitch tracks from sounds sung by subjects after listening to bells 1, 3, 4, 5, 7, 9, 10, 11, 12 of the Dumery carillon

Though our observations are limited in number, they seem to suggest that musically trained subjects tend to sing a note whose f_1 pitch corresponds to the hum of the bell sound rather than the prime or even the nominal. There are indications that composers like Hector Berlioz and Richard Strauss may also have identified the 'main pitch' of a (swinging) bell with the hum partial (cf. [80]). Further, there are empirical data from pitch perception experiments which corroborate that subjects matched a sine tone to the hum (f_1) of carillon bells [40]. The relative frequency of matches to the hum increased with f_1 of the bells (see [40], Fig. 19). This could be expected, to some degree, since the 16 carillon bells in this experiment were rather small (ranging from G_5 to D_7; f_1 for G5 is ca. 784 Hz). Experiments with sounds from swinging bells [56] showed that most subjects were able to identify the 'pitch category' (in terms of the semitones of a musical scale) of the strike note correctly while, in a number of bells and varying among the subjects, there was uncertainty whether the strike note would be located in the region of the prime or in the region of the hum partial. A possible explanation for this effect could be found in different patterns of spectral energy distribution in the sound recorded from various bells. As has been pointed out (above), spectral energy in sounds varies with the strength of excitation of bells as well as with directivity of radiation. Stronger excitation in general causes a higher spectral centroid and also an increase in spectral inharmonicity of the sound at the onset (see above).

Whether the strike note of a particular bell is perceived at the pitch of the prime or that of the hum (or still at another partial, or even in between partials; see [55, 51, 56]), appears to depend on several physical and also on psychoacoustic parameters. In about 30 % of the 137 (swinging) bells investigated by Terhardt and Seewann [56], more than one strike note could be determined algorithmically as well as in behavioural pitch matching experiments. Perception of more than one strike note for certain swinging bells was also observed in experiments with expert listeners [51, 55].

Taking the concept of Schouten [60], Schouten and t'Hart [61] and later findings one can argue that the strike note is a percept from a selection of strong partials in the bell sound which have nearly harmonic frequency ratios. These components form a complex that produces a virtual f_0 which, in general, induces a pitch close to that of the prime (f_2 in the bell sound) or the hum (f_1 in the bell sound). Hence, in regard to pitches perceived, in most cases $f_0 \simeq f_2$ or $f_0 \simeq f_1$ obtains. However, while the timbre of both the f_2 partial and the f_1 partial, if taken alone, appears rather soft, the timbre of the complex of partials giving rise to f_0 in general is found to be metallic. It is probably this timbral quality which distinguishes the strike note from the spectral pitches subjects identify with some of the more prominent partials, most of all, the prime, the minor third, the hum and the octave above the prime.

4.3 Tuning of the Dumery Carillon Bells in Regard to Prime (f_2) Frequencies

Accepting that, for many subjects listening to bells, the main pitch is either located at (or close to) the prime or at the hum, we calculated pitch frequencies with two different algorithms (AC, SHS) for 26 Dumery bell sounds at different time points after onset (t = 0.5, t = 1, and t = 2 s, and in several bell sounds also at t = 0.25″, t = 1.5″ and t = 3″). The data indicate that both algorithms in some sounds determined a pitch frequency close to the prime but in other sounds found a frequency close to the hum as the relevant pitch. Also, frequencies not directly related to either hum or prime as well as subharmonic frequencies (in general, a fraction of the hum frequency) were turned out by both AC and SHS pitch tracking algorithms. To illustrate the case, pitch tracks as calculated for bell no. 17 are shown in Fig. 20.

The AC pitch track (Fig. 20, solid line) starts at about 92 Hz and, after a jump to ca. 110 Hz, falls back to the initial frequency. To be sure, 92 Hz is about 1/3 of the frequency of the hum (273.35 Hz) and close to 1/6 of the prime (548.09 Hz) in this bell. The SHS track (dashed line) at t = 0.5″ and t = 1″ yields ca. 550 Hz and thus is very close to the prime. After a period of transition which reflects the effect of the rather fast decay of higher partials in the sound due to (viscous and acoustic) damping, the SHS pitch at t = 1.5″ and t = 2″ gives ca. 273 Hz, that is, the frequency of the hum note which has a long decay in the sound.

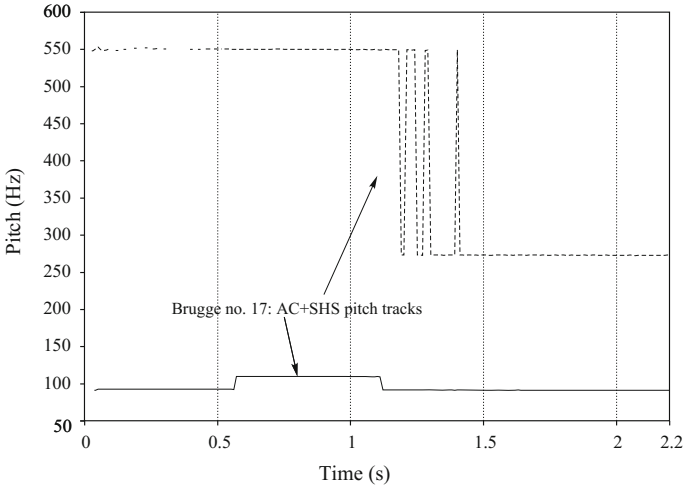


Fig. 20 Pitch tracks calculated for bell no. 17, AC and SHS algorithm

Assuming (a) that either the strike note or another main pitch of a bell can be expressed as a single frequency, and that (b) this frequency in many bell sounds is very close to that of the prime partial, we may devise the tuning of the 26 Dumery bells in terms of the prime frequencies measured from the bell sounds. Since some of the prime partials in fact are doublets, a decision has to be made whether to take the strongest component as relevant for determining the tuning (cf. [51]), or to use the mean of two adjacent frequency components as the respective pitch. In case the two components of a split partial differ significantly in amplitude level (A_1: $A_2 > 3$ dB), one may take the stronger component as perceptually relevant (the selection being justified by spectral masking). If, however, two components differ but little in level (A_1:$A_2 < 1.5$ dB), both can be part of the pitch percept (which may be somewhat blurred depending on the frequency distance of the two components as well as on their level relative to other partials). It seems justified, therefore, to calculate the mean of two frequency components for such doublets which have almost equal amplitudes. Hence, we have 26 frequencies for the prime partial to derive the tuning, which is given in cents for the intervals between these frequencies.

In regard to the question addressed above whether or not the Dumery carillon was intended to represent a quarter-comma meantone tuning, the data from Table 8 again offer some clues in this direction. First, there are whole tones smaller in size than 200 cent (and several relatively close to the meantone of 193 cent). Second, the intervals of the major third, the minor sixth and the major sixth in the scale based on the largest bell (G_2) are close to the meantone scale (see above; the major third and the major sixth are also close to just intervals). Third, the fourth is enlarged as in meantone tuning (the fifth G_2–D_3 in the first octave, however, is also enlarged while it should be diminished by about 5–6 cent in meantone tuning). Perhaps the strongest hint to a historical tuning (of which there were very many in use in the 17th and still in the 18th century; see e.g. [52, 54]) is that there are two clearly different semitone intervals, one representing the chromatic semitone (ideally with a frequency ratio of close to 25/24 $\sim$70.6 cent) and another representing the diatonic semitone 16/15 $\sim$111.7 cent. This difference is a characteristic of both meantone tuning and just intonation that has been eliminated in equal temperament (ET12) where all semitones have 100 cent. In sum, the tuning data of the 26 Dumery bells permits us to conclude that the model for the bell tuning might have been quarter-comma meantone temperament or one of the ‘well-tempered’ tunings in use between ca. 1680 and 1770 while in particular significant differences in the size of semitones in the scale indicate that ET12 cannot be considered as a template for the tuning of this carillon.

Table 8 Tuning of 26 Dumery carillon bells based on prime partial frequencies f_2

Bell	Prime f_2		Cents cum.	Bell	Prime f_2		Cents cum.
1 G_2	195.69		0	14 Bb_3	464.65		1498.1
		188.3				91.3	
2 A_2	218.04		188.3	15 B_3	489.82		1589.4
		194.8				115.3	
3 B_2	244.01		383.1	16 C_4	523.54		1704.7
		124.8				79.3	
4 C_3	262.25		507.9	17 $C\#_4$	548.09		1784.0
		71.7				110.6	
5 $C\#_3$	273.34		579.6	18 D_4	584.26		1894.6
		127.9				95.8	
6 D_3	294.30		707.5	19 Eb_4	617.50		1990.4
		67.4				110.3	
7 Eb_3	305.99		774.9	20 E_4	658.11		2100.7
		110.9				102.6	
8 E_3	326.24		885.8	21 F_4	698.30		2203.3
		116.0				66.5	
9 F_3	348.84		1001.8	22 $F\#_4$	725.65		2269.8
		77.6				130.2	
10 $F\#_3$	364.83		1079.4	23 G4	782.32		2400.0
		116.1				95.1	
11 G_3	390.13		1195.5	24 $G\#_4$	826.49		2495.1
		92.9				94.6	
12 $G\#_3$	411.64		1288.4	25 A_4	872.89		2589.7
		109.3				108.4	
13 A_3	438.46		1397.7	26 Bb_4	929.27		2698.1
		100.4					

5 Conclusion

The 26 Dumery bells (1742–48) from the carillon at Bruges belong to the rather small stock of historic bells from the 18th century that, from all we know, remained unaltered in regard to their shape (relevant for partial frequencies) and tuning to a musical scale until recently. Spectral analyses of the sound of these bells as well as other measurements revealed that each bell conforms closely to the pattern of 'principal partials' (hum, prime, third, fifth, octave, etc.) characteristic of a minor-third bell since about 1500 (G. de Wou). From what is known, Jacob van Eyck by about 1633 had identified at least some of the partials, and the famous Hemony brothers a few years later succeeded in founding carillon bells of highest quality that became kind of a standard in the Low Countries. The Hemony carillon bells are esteemed to this day as model specimens for the minor-third/octave bell.

Though Dumery lived and worked in the 18th century, he apparently was aware of acoustic and technical lore essential for casting high quality carillon bells. As our measurements demonstrate, deviations of the 26 bells from ideal norms and ratios are slight (a fact that is the more surprising if one considers the rather primitive conditions under which the bells were cast at Bruges, in the 1740s).

The tuning of the Dumery bells to a musical scale has at times been interpreted as close to quarter-comma meantone temperament. Our measurements with respect both to hum and prime partial frequencies show that there is some evidence in support of this interpretation; however, some of the data do not conform to this type of temperament. What can be concluded from the hum and prime partial frequencies (which, in most of the 26 bells, are close to forming a perfect octave) is that an unequal temperament is much more likely than ET12 for which actual deviations are too large. Whether the tuning followed one of the 'well-tempered' patterns (such as known from Werckmeister and his contemporaries) or was still aiming for one of the meantone temperaments (besides ¼ comma, there were several variants in practice), is difficult to decide.

Tuning of carillon bells cannot be investigated without regard to partial structure (addressed as 'inner harmony' of bells by campanologists), which in turn brings up the issue of the so-called strike note and its relevance for perception of the bell's pitch or, rather, pitches. While in the literature the position of the strike note as a virtual pitch often is located close to the prime partial frequency, our own measurements and experiments suggest that the hum partial frequency is also a candidate for the main pitch perceived from carillon bell sounds (the 'main pitch' apparently is perceived as 'implied' from quasi-harmonic segments in the spectral structure). As has been found in a number of previous experiments (and is confirmed by our own empirical data), most subjects perceive several spectral and virtual pitches when listening to sounds from minor-third bells. The multiplicity of concurrent pitches is the cause of pitch ambiguity. Moreover, the relative inharmonicity in minor-third bell spectra and the noisy transient sound shortly after the strike of the bell with a clapper hamper precise pitch estimates.

Acknowledgments Our recording of the Dumery carillon was greatly supported by the late Aimé Lombaert (1945–2008), who from 1984 until his untimely death was the chief carillonist (*stadsbeiaardier*) of the City of Bruges. We like to dedicate this study to his memory.

References

1. Augustus, B.: De Beiaard van Brugge, restauratie van de klokken, nieuwe klepels (restoration report). Asten: R. Eijsbouts (2011)
2. Lehr, A.: The system of the Hemony-carillon tuning. Acustica **3**, 101–104 (1951)
3. Schneider, A., Leman M.: Sonological and psychoacoustic characteristics of carillon bells. In: Leman M. (ed.) The Quality of Bells (= Proceedings of the 16th Meeting of the FWO Research Society). IPEM, Ghent (2002)

4. Slaymaker, F.H., Meeker, W.F.: Measurements of the tonal characteristics of carillon bells. J. Acoust. Soc. Am. **26**, 515–522 (1954)
5. Van Heuven, E.W.: Acoustical Measurements on church-bells and carillons. Diss. TH Delft (1949)
6. Fletcher, N., McGee, W., Tarnopolsky, A.: Bell clapper impact dynamics and the voicing of a carillon. J. Acoust. Soc. Am. **111**, 1437–1444 (2002)
7. Debut, V., Carvalho, M., Antunes, J.: An objective approach for assessing the tuning properties of historical carillons. In: Proceedings of the Stockholm Musical Acoustics Conference 2013 (SMAC 2013), pp. 549–556. Logos-Verlag, Berlin (2013)
8. Lehr, A., Truyen, W., Huybens, G.: The Art of the Carillon in the Low Countries. Lannoo, Tielt (1991)
9. Augustus, L.M.: Personal communication to M. Leman (including the 1969 restoration report) (1998)
10. Rossing, Th: Science of Percussion Instruments. World Scientific, Singapore (1993)
11. Ellerhorst, W., Klaus, G.: Handbuch der Glockenkunde. Verlag der Martinus-Buchhandlung, Weingarten (1957)
12. Lehr, A.: The geometrical limits of the carillon bell. Acustica **86**, 543–549 (2000a)
13. Schad, C.R., Frik, G.: Klangfiguren einer Glocke. Acustica **78**, 46–54 (1993a)
14. Schad, C.R., Frik, G.: Zur Klangstruktur einer berühmten Glocke. Acustica **78**, 110–114 (1993b)
15. Van Baak Griffoen, R.: Jacob van Eyck's Der Fluyten Lust-Hof (1644-c. 1655). Vereniging voor Nederlandse Muziekgeschiedenis, Den Haag (1991)
16. Van den Hul, D.: Klokkenkunst te Utrecht tot 1700. Ph.D. dissertation, University of Utrecht (1982)
17. Beeckman, I.: Journal tenu par Isaac Beeckman. In: de Waard C. (eds.) T. III, La Haye: Nijhoff (1945)
18. Leman, M.: The 18th century Dumery-carillon of Brugge: Manuscripts at the Central Library of Ghent University. In: Leman M. (ed.) The quality of bells. (= Proceedings of the 16th Meeting of the FWO Research Society). IPEM, Ghent (2002)
19. Fletcher, N., Rossing, T.: The Physics of Musical Instruments. Springer, New York (2nd edn. 1998) (1991)
20. Bader, R.: Additional modes in a Balinese gender plate due to its trapezoid shape. In: Bader R., Neuhaus C., Morgenstern U. (eds). Concepts, Experiments, and Fieldwork: Studies in Systematic Musicology and Ethnomusicology, pp. 95–111. P. Lang, Frankfurt/M (2010)
21. Flügge, W.: Statik und Dynamik der Schalen, 3rd edn. Springer, Berlin (1962)
22. Göldner, H.: Lehrbuch höhere Festigkeitslehre, Bd 2, 2nd edn. Fachbuchverlag, Leipzig (1989)
23. Graff, K.: Wave Motion in Elastic Solids. Clarendon Press, Ohio State University Press, Oxford/Columbus (Reprint Dover, New York, 1991) (1975)
24. Grützmacher, M., Kallenbach, W., Nellesen, E.: Akustische Untersuchungen an Kirchenglocken. Acustica 16, 34–45 (1965/66)
25. Grützmacher, M., Kallenbach, W., Nellessen, E.: Vergleich der nach verschiedenen Verfahren berechneten Eigenfrequenzen kreiszylindrischer Schalen mit gemessenen Werten. Acustica **17**, 79–89 (1967)
26. Junger, M., Feit, D.: Sound, structures, and their interaction. Acoust. Soc. of Am. Woodbury, N.Y.: (= Reprint of the 2nd edn. MIT Press, Cambridge, MA 1986) (1993)
27. Leissa, A.: Vibration of shells. Acoust. Soc. Am. Woodbury, NY (= Reprint of the 1973 NASA ed.) (1993b)
28. Rayleigh, J.: The Theory of Sound, vol. 1. Macmillan, London (reprint Dover, New York) (1877/1945)
29. Chladni, E.F.F.: Die Akustik. Breitkopf & Haertel, Leipzig (2nd edn. 1830) (1802/1830)
30. Cremer, L., Heckl, M.: Körperschall. Springer, Berlin (1967) (3rd revised edition Cremer, L., Heckl, M., Petersson, B.: Structure-borne sound: structural vibrations and sound radiation at audio frequencies. Springer, Berlin (2010))

31. Leissa, A.: Vibration of Plates. Acoust. Soc. Am. Woodbury, N.Y. (= Reprint of the 1973 NASA ed.) (1993a)
32. Schneider, A.: Tonhöhe – Skala – Klang. Akustische, tonometrische und psychoakustische Studien auf vergleichender Grundlage. Orpheus-Verlag für syst. Musikwiss, Bonn (1997)
33. Schad, C.R.: Werkstoffeinflüsse auf die Klangeigenschaften von Glocken. In: Glocken in Geschichte und Gegenwart. Beiträge zur Glockenkunde [Bd 1], bearbeitet von Kurt Kramer, pp. 148–157. Badenia-Verlag, Karlsruhe (1986)
34. Lau, B., Bader, R., Schneider, A., Wriggers, P.: Finite element transient calculation of a bell struck by its clapper. In: Bader R., Neuhaus C., Morgenstern U. (eds.) Concepts, Experiments, and Fieldwork: Studies in Systematic Musicology and Ethnomusicology, pp. 137–156. P. Lang, Frankfurt/M
35. Perrin, R., Charnley, T.: Group theory and the bell. J. Sound Vibr. **31**, 411–418 (1973)
36. Gehrtsen, Chr, Vogel, H.: Physik, 17th edn. Springer, Berlin (1993)
37. Euler, L.: Tentamen de sono campanorum (1776). In: Leonardi Euleri…Opera Omnia, Series II, vol. X. Orell-Füssli, Bern/Lausanne (1776/1957)
38. Chung, J., Lee, J.: Vibration analysis of a nearly axisymmetric shell structure using a new finite ring element. J. Sound Vibr. **219**, 35–50 (1999)
39. Charnley, T., Perrin, R.: Torsional vibrations of bells. J. Sound Vibr. **40**, 227–231 (1975)
40. Fleischer, H.: Schwingung und Tonhöhe von Glockenspielglocken. (= Seminarberichte aus dem Gebiet Technische Mechanik und Flächentragwerke 1/96). München: Univ. der Bundeswehr (1996)
41. Lehr, A.: Partial groups in the bell sound. J. Acoust. Soc. Am. **79**, 2000–2011 (1986)
42. Perrin, R., Charnley, T., DePont, J.: Normal modes of the modern English church bell. J. Sound Vibr. **90**, 29–49 (1983)
43. Perrin, R., Charnley, T., Banu, H., Rossing, Th: Chladni‘s law and the modern English church bell. J. Sound Vibr. **102**, 11–19 (1985)
44. Rossing, Th, Sathoff, H.J.: Modes of vibration and sound radiation from tuned handbells. J. Acoust. Soc. Am. **68**, 1600–1607 (1980)
45. Rossing, T., Perrin, R.: Vibration of bells. Appl. Acoust. **20**, 41–70 (1987)
46. Schad, C.R.: Form-Klang-Rechnungen an Glocken. Acustica **64**, 272–285 (1987)
47. Schoofs, A.J.: Computer aided bell design and optimization. In: Leman M. (ed.) The Quality of Bells (= Proceedings of the 16th Meeting of the FWO Research Society). Ghent: IPEM
48. Lehr, A.: A statistical investigation of historical swinging bells in West Europe. Acustica **74**, 97–108 (1991)
49. Lehr, A., et al.: A carillon of major-third bells. Music Percept. **4**, 243–280 (1987)
50. Lehr, A.: The removal of warble or beats in the sound of a bell. Acustica **86**, 550–556 (2000b)
51. Hibbert, W.A.: The Quantification of strike pitch and pitch shifts in church bells. Ph.D. dissertation, Milton Keynes, The Open University, UK (2008)
52. Barbieri, P.: Enharmonic Instruments and Music 1470–1900. Levante, Latina (2008)
53. Hall, D.: Musical Acoustics. Wadsworth, Belmont (1980)
54. Lindley, M.: Stimmung und Temperatur. In Hören, Messen und Rechnen in der frühen Neuzeit (Geschichte der Musiktheorie, Bd 6, ed by F. Zaminer), pp. 109–331. Wiss. Buchges, Darmstadt (1987)
55. Bruhn, G.: Über die Hörbarkeit von Glockenschlagtönen. Bosse, Regensburg (1980)
56. Terhardt, E., Seewann, M.: Auditive und objektive Bestimmung der Schlagtonhöhe von historischen Kirchenglocken. Acustica **54**, 129–144 (1984)
57. de Cheveigné, A.: Pitch perception models. In: Plack C., Oxenham J., Fay R., Popper A. (eds.) Pitch. Neural Coding and Perception, pp. 169–233. Springer, New York (2005)
58. Biehle, J.: Die Analyse des Glockenklanges. Archiv für Musikwiss **1**, 289–312 (1919)
59. Terhardt, E.: Akustische Kommunikation. Springer, Berlin (1998)
60. Schouten, J.F.: The Perception of pitch. Philips Tech. Rev. **5**, 286–294 (1940)
61. Schouten, J.F., t’Hart, J.: De Slagtoon van Klokken. In *Nederlands Akoestisch Genootschap*, Delft, Publikatie nr. vol. 7, pp. 8–19. (Engl. trans. in Rossing 1984) (1965)

62. de Boer, E.: On the «Residue» and auditory pitch perception. In: Keidel W., Neff W. (eds.) Handbook of Sensory Physiology, vol. V,3. Springer, Berlin (1976) (ch. 13)
63. Plomp, R.: Aspects of tone sensation. Academic Press, London (1976)
64. Houtsma, A.: Pitch perception. In: Moore B. (eds.) Hearing, pp. 267–295. Academic Press, San Diego (1995)
65. Moore, B.: Introduction to the Psychology of Hearing, 5th edn. Academic Press, Amsterdam (2007)
66. Schneider, A.: Psychophysics and psychoacoustics. In: Bader R. (ed.) Handbook of Systematic Musicology. Springer, Berlin (in press)
67. Eggen, J., Houtsma, A.: The pitch perception of bell sounds. Instituut voor Perceptie Onderzoek (Eindhoven, Nl). Annual Report vol. 21, pp. 15–23 (1986)
68. Boersma, P.: Accurate short-term analysis of the fundamental frequency and the harmonic-to-noise ratio of a sampled sound. In: Proceedings, Institute of Phonetic Sciences, University of Amsterdam, vol. 17, pp. 97–110 (1993)
69. Tolonen, T., Karjalainen, M.: A computationally efficient multipitch analysis model. IEEE Trans. Speech Audio Process. **8**, 708–716 (2000)
70. Hermes, D.: Measurement of pitch by subharmonic summation. J. Acoust. Soc. Am. **83**, 257–264 (1988)
71. Zwicker, E., Fastl, H.: Psychoacoustics. Facts and Models, 2nd edn. Springer, Berlin (1999)
72. Beauchamp, J.: Analysis and synthesis of musical instrument sounds. In: Beauchamp, J. (ed.) analysis, Synthesis, and Perception of Musical Sounds, pp. 1–89. Springer, New York (2007)
73. von Helmholtz, H.: Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik, 3rd edn. Vieweg, Braunschweig (1870)
74. Hartmann, W.: Signals, Sound, and Sensation. Springer, New York (1998)
75. Schneider, A.: Inharmonic sounds. Implications as to 'pitch', 'timbre', and 'consonance'. J. New Music Res. **29**, 275–301 (2000)
76. Schneider, A., Frieler, K.: Perception of harmonic and inharmonic sounds: results from ear models. In: Ystad S., Kronland-Martinet R., Jensen K. (eds.) Computer Music Modeling and Retrieval. Genesis of Meaning in Sound and Music, pp. 18–44. Springer, Berlin (2009)
77. Schneider, A.: Complex inharmonic sounds, perceptual ambiguity, and musical imagery. In: Godøy, R.I., Jørgensen, H. (eds.) Musical Imagery, pp. 95–116. Swets & Zeitlinger, Lisse, Abingdon (2001)
78. Bossin, J.: Musik für Carillon 1600–1900. Die Suche nach einem verschollenen Repertoire. In: Glocken in Geschichte und Gegenwart. Beiträge zur Glockenkunde, Bd 2, bearbeitet von K. Kramer. Karlsruhe: Badenia-Verlag, pp. 113–129 (1995)
79. Kramer, K., Menschick, W., Wagner, G.: Zur Benennung der Glockentöne. In: Glocken in Geschichte und Gegenwart. Beiträge zur Glockenkunde, Bd 2, bearbeitet von K. Kramer, pp. 62–67. Badenia-Verlag, Karlsruhe (1995)
80. Wagner, G.: Glockenklang und Harmonie. In: Glocken in Geschichte und Gegenwart. Beiträge zur Glockenkunde [Bd 1], bearbeitet von Kurt Kramer, pp. 73–78. Badenia-Verlag, Karlsruhe (1986)

Authors Biography

Albrecht Schneider (b. 1949) worked as a professor of systematic musicology in the Institute of Musicology, University of Hamburg, from 1981 to 2011 (and as an emeritus since). He has taught as visiting professor in the Dept. of Ethnomusicology and Systematic Musicology of UCLA and has lectured also in various European institutions. He is the author, co-author, editor and co-editor of a number of books and has contributed chapters and articles to scientific handbooks, encyclopedias, and journals (for bio-bibliographical data, see R. Bader, Chr. Neuhaus, and U.

Morgentern [eds]. *Concepts, Experiments, and Fieldwork: Studies in Systematic Musicology and Ethnomusicology*. Frankfurt/M., Berne, Brussels: P. Lang 2010, and www.fbkultur.uni-hamburg.de/sm/personen.html).

Marc Leman is Methusalem research professor in systematic musicology and director of IPEM, the research center of the musicology department at Ghent University. He holds MA degrees in musicology and philosophy, and did his PhD on computer modeling of tonality perception. He published more than 350 articles, and several books among which the monographs "Embodied music cognition and mediation technology", MIT Press, 2007 and "The expressive moment: how interaction (with music) shapes musical empowerment", MIT Press, 2016; and the co-edited book "Musical gestures: sound, movement, and meaning", Routledge, 2010. His lab is an international meeting place for researchers working on expressive interactions with music, using embodiment and action theory as a point of departure. In 2007 he became professor of the Methusalem, renewed in 2015. In 2014 he was holder of the Franqui chair at the Université de Mons. In 2015 he became laureate of the FWO (Flemish Fund for Scientific Research) Excellence Award Ernest-John Solvay for Language, Culture and Social Sciences.

Source Width in Music Production. Methods in Stereo, Ambisonics, and Wave Field Synthesis

Tim Ziemer

Abstract Source width of musical instruments, measured in degrees, is a matter of source extent and the distance of the observer. In contrast to that, perceived source width is a matter of psychological organization of sound. It is influenced by the sound radiation characteristics of the source and by the room acoustics and restricted by masking and by localization accuracy. In this chapter perceived source width in psychoacoustics and apparent source width in room acoustical research are revisited. Source width in music recording and production practice in stereo and surround as well as in ambisonics and wave field synthesis are addressed. After the review of the literature an investigation is introduced. The radiation characteristics of 10 musical instruments are measured at 128 angles and the radiated sound is propagated to potential listening positions at 3 different distances. Here, monaural and binaural sound quantities are calculated. By means of multiple linear regression, the physical source extent is predicted by sound field quantities. The combination of weighted interaural phase differences in the sensitive frequency region together with the number of partials in the quasi-stationary part of instrumental sounds shows significant correlation with the actual source extent of musical instruments. The results indicate that these parameters might have a relevant effect on perceived source extent as well. Consequently, acoustic control over these parameters will increase psychoacoustic control concerning perceived source extent in audio systems.

1 Introduction

Due to extensive and well-elaborated investigations in the field of psychoacoustics and subjective room acoustics within the last hundred years, a lot of knowledge about the auditory perception of source extent has been acquired. It is outlined in

T. Ziemer (✉)
Institute of Systematic Musicology, University of Hamburg,
Neue Rabenstr. 13, 20354 Hamburg, Germany
e-mail: tim.ziemer@uni-hamburg.de

A. Schneider (ed.), *Studies in Musical Acoustics and Psychoacoustics*,
Current Research in Systematic Musicology 4,
DOI 10.1007/978-3-319-47292-8_10

the following section. In music recording, mixing and mastering practice, several methods to control the perceived source extent have been established for channel based audio systems like stereo and surround. More recently, novel approaches for object based audio systems like ambisonics and wave field synthesis have been proposed. These are revisited and examined from a psychoacoustic point of view. Following this theoretic background, an investigation to illuminate the direct relationship between source width and signals reaching the ears is presented. For this task, the radiation characteristics of 10 acoustical instruments are recorded. By means of a simplification model, ear signals for 384 listening positions are calculated, neglecting room acoustical influences. Then, physical measures derived from the field of psychoacoustics and subjective room acoustics, are adapted to an anechoic environment. From these measures the actual source extent is predicted. Assuming that the perceived and the actual physical source extent largely coincide, these predictors give clues about the ear signals necessary to create the impression of a certain source width. This knowledge can be utilized for control over apparent source width in audio systems by considering the ear signals, instead of channel signals. It is an attempt at answering the question how perceived source extent is related to physical sound field quantities. A preliminary state of this study has been presented in Ziemer [50].

2 Perception of Source Width

Spatial hearing has been investigated extensively by researchers both in the field of psychoacoustics and in subjective room acoustics. Researchers in the first area tend to make listening tests under controlled laboratory conditions with artificial stimuli, such as clicks, noise and Gaussian tones. They investigate localization and the *perception of source width*. Researchers from the field of subjective room acoustics try to find correlations between sound field quantities in room impulse responses and sound quality judgments reported by expert listeners. Alternatively, they present artificial room acoustics to listeners, i.e. they use loudspeaker arrays in anechoic chambers. They observed that reflections can create the impression of a source that sounds even wider than the physical source extent. This auditory impression is referred to as *apparent source width*. Results from both research fields are addressed successively in this section.

2.1 Perceived Source Width in Psychoacoustics

Spatial hearing has been investigated mostly with a focus on sound source localization. Blauert [6] is one of the most comprehensive books about that topic. The localization precision lies around 1° in the frontal region, with a localization blur of about ±3.6°. Localization cues are contained in the head-related transfer function

(HRTF). It describes how a sound signal changes from the source to the ears. Monaural cues like overall volume and the distribution of spectral energy mainly serve for distance hearing. The further the source, the lower the volume. Due to stronger attenuation of high frequencies in the air, distant sources sound more dull than proximate sources. Furthermore, low frequencies from behind easily diffract around the pinnae. For high frequencies, the pinnae create a wave shadow. So the spectral energy distribution also helps for localization in the median plane. Binaural cues are interaural time differences (ITD) and interaural level differences (ILD) of spectral components. In dichotic playback, interaural phase differences (IPD) can be created without introducing ITD. Using forced-choice listening tasks and magnetoencephalography, Ross et al. [42] could prove, both behavioristically and neurally, that the human auditory system is sensitive to IPD below about 1.2 kHz.

Blauerte considers the localization blur the just noticeable difference (JND) in location whereas Zwicker and Fast1 consider it as precision with which the location of one stationary sound source can be given.[1] Both interpretations allow to hypothesize that the localization blur is related to width perception. The inability to name one specific angle as source angle may be due to the perception of a source that is extended over several degrees.

It is clear, however, that source localization and the perception of source width are not exactly the same. Evidence for this is the precedence effect which is sometimes referred to as *Haas effect* or *law of the first wavefront*.[2] The first arriving wave front is crucial for localization. Later arriving reflections hardly affect localization but can have a strong influence on the perceived source extent. Only a few authors investigated perceived source extent of the direct sound in absence of reflections. Hirvonen and Pulkki [24] have investigated the perceived center and spatial extent under anechoic conditions with a 45°-wide loudspeaker array consisting of 9 speakers. Through these, one to three non-overlapping, consecutive narrow-band noises were played by each speaker. The signals arrive simultaneously at a sweet-spot to minimize ITD and bias that results from the precedence effect. All loudspeakers were active in all runs. In all cases the perceived width was less than half the actual extent of the loudspeaker array. The authors were not able to predict the perceived width from the distribution of signals over the loudspeaker array. Investigating the relationship between perceived source width and ear signals, instead of loudspeaker signals, might have disclosed quantitative relationships. Furthermore, it might be difficult for a subject to judge the width of a distributed series of noise because such a signal is unnatural and not associated to a known source or a previously experienced listening situation. Natural sounds may have led to more reliable and predictable results. However, based on their analysis of

[1]Cf. Blauert [6], pp. 37f and Zwicker and Fastl [56], p. 309.

[2]See e.g. Haas [21], Blauert and Cobben [8].

channel signals they can make the qualitative statement that the utilized frequency range seems to have a strong impact on width perception.[3]

Potard and Burnett [39] found that "shapes", i.e. constellations of active loudspeakers, could be discriminated in the frontal region in cases of decorrelated white noise and 3 kHz high-pass noise in 42.5 and 41.4 % of all cases. Neither were subjects able to perform this task with 1 kHz low-pass noise and blues guitar, nor were they able to discriminate shapes in the rear for any kind of tested signal. The authors point out that perception of width and identification of source shape are highly dependent on the nature of the source signal. Furthermore, they observed that 70.4 % of all subjects rated a set of decorrelated sources more natural than a single loudspeaker for naturally large auditory events like crowd, beach etc. The findings that shapes of high-pass noise were discriminated better than shapes of low-pass noise underlines the importance of high-frequency content for the recognition of shapes. It could mean that ILD play a crucial role for the recognition of shapes. ILD mainly occur at high frequencies whereas low-pass noise mainly created IPD. The fact that high pass noise was discriminated better than blues guitar could furthermore denote that continuous sounds contain more evaluable information than impulsive sounds. The observation that only shapes in the frontal region could be discriminated may imply that experience with visual feedback improves the ability to identify constellations of sound sources. However, these assumptions are highly speculative and need to be confirmed by further investigations.

These two experiments demonstrate that subjects fail to recognize source width or shapes of unnaturally radiating sources, i.e. loudspeakers. Furthermore, mostly unnatural sounds are used, i.e. sounds that are not associated to a physical body, in contrast to the sound of musical instruments. In these two investigations loudspeaker signals are controlled. Control over the sound that actually reaches the listeners' ears might reveal direct cues concerning the relationship between the sound field and the perceived source width. Like blauerte states: "The sound signals in the ear canals (ear input signals) are the most important input signals to the subject for spatial hearing."[4] The investigation presented in Sect. 4 follows this paradigm, not controlling source signals but investigating what actually reaches the listeners' ears. The source signals are notes, played on real musical instruments including their natural sound radiation characteristics. Such signals are well-known to human listeners and associated with the physical extent of the instrument.

In many situations in which the listener is far away from the source, the physical source width is less than the localization blur. This is the case for most seats in concert halls for symphony music and opera. Here, the room acoustics, i.e. reflections, play a larger role for the auditory perception of source extent than the

[3]The complete investigation is documented in Hirvonen and Pulkki [24]. Contrary to width, they succeeded to replicate perceived source center by different adaptations of Raatgever's frequency weighting function.

[4]Blauert [6], p. 51.

direct sound. On the other hand, the radiation characteristics of sound sources have an immense influence on the room response. Apparent source width in room acoustics is discussed in the following.

2.2 Apparent Source Width in Room Acoustics

In the context of concert hall acoustics many investigations have been carried out to find relationships between physical sound field parameters and (inter-)subjective judgments about perceived source extent or overall sound quality. Since our acoustic memory is very short,[5] a direct comparison between listening experiences in different concert halls is hardly possible. Hence, listening tests have been conducted with experts, like conductors and music critics, who have long-term experience with different concert halls. Another method is to present artificially created and systematically altered sound fields or even auralize the complete room acoustics of concert halls. An overview about subjective room acoustics can be found in Beranek [4] and Gade [18].

In the context of subjective room acoustics, the apparent source width (ASW) is often defined as the auditory broadening of the sound source beyond its optical size.[6] Most authors agree that ASW is especially affected by direct sound and early reflections, arriving within the first 50–80 ms. Other terms that are used to describe this perception are *image* or *source broadening*, *subjective diffuseness* or *sound image spaciousness*.[7] All these terms are treated as the same in this chapter. The term *perceived source extent* is used to describe the auditory perception regardless of the quantities or circumstances that cause this impression.

The early lateral energy fraction ($\mathrm{LEF_{E4}}$) is proposed as ASW measure in international standards. It describes the ratio of lateral energy to the total energy at a receiver position $\mathbf{r}$ like[8]

$$\mathrm{LEF_{E4}}(\mathbf{r}) = \frac{\int_{t=5\,\mathrm{ms}}^{80\,\mathrm{ms}} p_8^2(\mathbf{r},t)\mathrm{d}t}{\int_{t=0}^{80\,\mathrm{ms}} p^2(\mathbf{r},t)\mathrm{d}t} \,. \tag{1}$$

Here, $p^2(\mathbf{r},t)$ is the squared room impulse response, measured by an omnidirectional microphone. The function $p_8^2(\mathbf{r},t)$ is the squared recording by a figure-of-eight-microphone whose neutral axis points towards the source. The subscript E stands for "early" and includes the first 80 ms. The subscript 4 denotes that the four octave bands around 125, 250, 500 and 1000 Hz are considered.

[5] See e.g. Gade [18], p. 304.

[6] See e.g. Blau [5], p. 720.

[7] See e.g. Yanagawa and Tohyama [47] and Yanagawa et al. [48].

[8] See e.g. Deutsches Institut f˙ur Normung [15], pp. 20f and Beranek [4], pp. 519 and 161.

The figure-of-eight microphone mainly records lateral sound whereas signals from the median plane largely cancel out. Hence, $\mathrm{LEF_{E4}}$ is the ratio of lateral to median sound or signal difference to signal coherence. The larger the value, the wider the expected ASW. In a completely diffuse field a value of $\mathrm{LEF_{E4}} = 0.33$ would occur.[9]

Beranek [4] found a significant negative correlation between ASW and the early interaural crosscorrelation ($\mathrm{IACC_{E3}}$). The subscript 3 denotes that the mean value of three octave bands around 500, 1000 and 2000 Hz is considered. $1 - \mathrm{IACC_{E3}}$ is also known as *binaural quality index* (BQI). BQI shows positive correlation to ASW. It is calculated from the $\mathrm{IACC_E}$, which is the maximum absolute value of the interaural crosscorrelation function (IACF) as measured from band passed portions of impulse response recordings with a dummy head:

$$\mathrm{IACF_E}(\mathbf{r}, \tau) = \frac{\int_{t=0}^{80\,\mathrm{ms}} p_\mathrm{L}(\mathbf{r}, t) p_\mathrm{R}(\mathbf{r}, t+\tau)\mathrm{d}t}{\sqrt{\int_{t=0}^{80\,\mathrm{ms}} p_\mathrm{L}^2(\mathbf{r}, t)dt \int_{t=0}^{80\,\mathrm{ms}} p_\mathrm{R}^2(\mathbf{r}, t)\mathrm{d}t}} \tag{2}$$

$$\mathrm{IACC_E}(\mathbf{r}) = \max|\mathrm{IACF_E}(\mathbf{r}, \tau)| \tag{3}$$

$$\mathrm{BQI}(\mathbf{r}) = 1 - \mathrm{IACC_{E3}}(\mathbf{r}) \tag{4}$$

The subscripts L and R denote the left and the right ear. The variable τ describes the time lag, i.e. the interval in which the interaural cross correlation is searched; $\tau \in (-1, 1)$ ms roughly corresponds to the ITD of a completely lateral sound. The IACC is calculated individually for each of the three octave bands. Their mean value is $\mathrm{IACC_{E3}}$. Beranek [4] found a reasonable correlation between LEF and BQI, which is not confirmed by other authors.[10] Ando even found neural correlates to BQI in the brainstem of the right hemisphere which is a strong evidence that the correlation of ear signals is actually coded and processed further by the auditory system.[11] It is conspicuous that two predictors of ASW—namely $\mathrm{LEF_{E4}}$ and BQI—consider different frequency regions. In electronically reproduced sound fields Okano et al. [37] have found that a higher correlation could be achieved when combining BQI with $G_\mathrm{E,low}$, the average early strength of the 125- and 250 Hz-octave band which is defined as

$$G_\mathrm{E,low}(\mathbf{r}) = 10\,\mathrm{lg}\frac{\int_{t=0}^{80\,\mathrm{ms}} p^2(\mathbf{r}, t)\mathrm{d}t}{\int_{t=0}^{\mathrm{dir}} p_\mathrm{ref}^2(t)\mathrm{d}t}\ . \tag{5}$$

$G_\mathrm{E,low}$ is the ratio between sound intensity of a reverberant sound and the pure direct sound p_ref. lg is the logarithm to the base 10 and the denominator represents

[9] According to Gade [18], p. 309.

[10] Cf. Beranek [4], p. 528 versus Blau [5] and Gade [18], p. 310.

[11] See Ando [2], p. 5.

the integrated squared sound pressure of the pure direct sound, which is proportional to the contained energy. The finding that strong bass gives rise to a large ASW even when creating coherent ear signals is not surprising. In nature only rather large sources tend to radiate low-frequency sounds to the far field. Here, the wavelengths are so large that barely any interaural phase- or amplitude differences occur. From psychoacoustic investigations it is known that monaural cues help for distance hearing. And distance, of course, strongly affects source width if we consider the relative width in degrees from a listener's point of view.

An alternative measure that includes the enlarging effect of strong bass frequencies is the *interaural difference*

$$\mathrm{IAD}(\mathbf{r}) = 10\,\mathrm{lg}\left(\frac{\mathrm{eq}(p_\mathrm{L}(\mathbf{r},t) - p_\mathrm{R}(\mathbf{r},t))^2}{p_\mathrm{L}^2(\mathbf{r},t) + p_\mathrm{R}^2(\mathbf{r},t)}\right). \tag{6}$$

This measure is proposed in Griesinger [20]. Basically, it is the difference signal of the squared dummy head recordings divided by the sum of their squared signals. The signal difference between the two dummy head ears is similar to a recording with a figure-of-eight microphone, and quantifies lateral sound energy. Their sum approximate an omnidirectional recording. Here, phase inversions cancel out and the mono component of the sound field is quantified. The factor eq stands for an equalization of the difference signal. Frequencies below 300 Hz are emphasized by 3 dB per octave. Due to their large wavelengths, bass frequencies hardly create interaural phase differences, even in a reverberant sound field. Consequently, a strong bass reduces values for $\mathrm{LEF}_{\mathrm{E4}}$, which contradicts the listening experience. This is probably the reason why the BQI does not consider such low frequencies. The equalization in the IAD counteracts this false trend. Unfortunately, the paper does not report any experience with this measure and its relationship to ASW.

Another approach to take the widening effect of low frequencies into account is to consider the width of the major IACF peak (W_{IACC}). Low frequencies tend to create wide IACF peaks, because small time lags barely affect phase. So W_{IACC} is related to the distribution of spectral energy. Shimokura et al. [44] even states that W_{IACC} is correlated to the spectral centroid of a signal. In Ando [2], it is described that a combination like

$$\mathrm{ASW}_{\mathrm{pre}} = \alpha(\mathrm{IACC})^{3/2} + \beta(W_{\mathrm{IACC}})^{1/2} \tag{7}$$

yields a very good prediction of ASW of band pass noise, if α and β are calculated for individuals.[12] For multi-band noise, the binaural listening level (LL) is an important additional factor.

Of all objective parameters that are commonly measured in room acoustical investigations, the $\mathrm{IACC}_{\mathrm{E}}$, and the strength G belong to the quantities that are most sensitive to variations of the sound radiation characteristics. In Martin et al. [35],

[12]See Ando [2], p. 130ff.

acoustical parameters are measured for a one source-receiver constellation but with two different dodecahedron loudspeakers. Although both loudspeakers approximate an omnidirectional source, deviations of G and BQI are larger than the just noticeable difference, i.e. they are assumed to be audible. In their experiment, this is not the case for $\mathrm{LEF_{E4}}$. This is probably the case because $\mathrm{LEF_{E4}}$ mainly considers low frequencies. Dodecahedron loudspeakers approximate an omnidirectional source much better at low frequencies than at high frequencies. Although good correlations between reported ASW and measured BQI could be found in many studies, this measure is not always a reliable predictor. It has been found that BQI tends to have massive fluctuation even when only slightly moving the dummy head. The same is true for $\mathrm{LEF_{E4}}$. These fluctuations are not in accordance with listening experiences.[13] When sitting in one concert hall seat and slightly moving the head, the ASW does not change as much as the BQI and the $\mathrm{LEF_{E4}}$ indicate. From a perceptual point of view, an averaging of octave bands is questionable, anyway. The auditory system rather averages over critical bands which can be approximated better by third-octave bands. Consequently, these measures are not valid for one discrete listening position $\mathbf{r}$. Their spatial averages over many seats rather give a good value for the overall width impression in the concert hall under consideration. This finding has been confirmed partly in Blau [5]. In listening tests with synthetic sound fields, the author could not find an exploitable correlation between ASW and BQI when considering all investigated combinations of direct sound and reflection. Only after eliminating individual combinations a correlation could be observed. He could prove that the fluctuations of BQI over small spatial intervals is not the only reason for the low correlation. He observed a higher correlation between ASW and $\mathrm{LEF_{E4}}$, which could explain $R^2 = 64$ % of the variance with one pair of reflections and $R^2 = 88$ % with multiple reflections. Assuming that frequencies above 1 kHz as well as the delay of single reflections may play a considerable role, Blau [5] proposed

$$\mathrm{RL_E} = 10\,\mathrm{lg}\frac{\sum_{i=1}^{n} a_i \sin\alpha_i E_i}{E_D + \sum_{i=1}^{n}(1 - a_i \sin\alpha_i)E_i} \tag{8}$$

as measure for ASW.[14] Here, i is the time window index. Time windows have a length of 2 ms and an overlap of at least 50 %. The upper bound n is the time window that ends at 80 ms. The weighting factor $a_i = 1 - e^{-t_i/15\,\mathrm{ms}}$ is an exponentially growing factor to emphasize reflections with a larger delay. α_i is the dominant sound incidence angle in the ith time window. It is estimated from an IACF of the low-passed signals weighted by a measure of ILD. E_D is the energy of the direct sound, E_i is the reflected energy contained in the ith time window.

[13]For details on the spatial fluctuations of BQI and LEFE4 refer to de Vries et al. [14].

[14]See Blau [5], p. 721.

The RL_E explained 89–91 % of the variance. It could be proved that the BQI changes when exciting the room using continuous signals instead of an impulse.[15] This finding may indicate that this measure cannot be applied to arbitrary signals. On the other hand, Potard and Burnett [39] already found out that the discrimination of shapes works with continuous high-pass noise but not with blues guitar. Likewise, width perception could be different for impulsive and continuous signals, so a measure for ASW does not necessarily need to have the same value for an impulse and a continuous signal. In the end, the BQI does not claim to predict ASW under conditions other than concert hall acoustics. It considers an omnidirectional impulse and does neither make a clear separation between direct sound and reflections nor does it take the radiation characteristics of sources into account. The radiation characteristics have a strong influence on the direct sound and the room acoustical response.

In Shimokura et al. [44], the IACC of a binaural room impulse response is differentiated from an $IACC_{SR}$ of an arbitrary source signal. They propose some methods to translate $IACC_{SR}$ to IACC, which are out of scope of this chapter. The authors convolve dry signals of musical instruments with binaural room impulse responses to investigate the relationship between perceived width and $IACC_{SR}$ with different signals. This way, different performances in the same hall can be compared as well as the same performance in different halls. By multiple linear regression the authors tried to predict reported diffuseness (SV) from descriptors of the signals' autocorrelation functions (ACFs) by

$$\mathrm{SV}(\mathbf{r}) = a\mathrm{IACC}(\mathbf{r}) + b\tau_e + cW_{\phi(0)}(\mathbf{r}) + d\ . \tag{9}$$

Here, $W_{\phi(0)}$ is the width of the first IACF peak and τ_e is the duration until the envelope of the ACF falls by 10 dB. It is 0 for white noise and increases when decreasing the bandwidth and converges towards ∞ for a pure tone. The contribution of IACC was significant for eight of nine subjects, whereas the contribution of τ_e and $W_{\phi(0)}$ was only significant for four and two of nine. Consequently, the multiple linear regression failed to explain SV of all subjects. Just as in the approach of Ando [2], Eq. 7, the factors a, b and c had to be adjusted for each individual. Shimokura et al. [44] observed that W_{IACC} was only significant for one individual subject which contradicts the findings of Ando [2]. Both approaches explain subjective ratings on the basis of objective parameters but their findings do not exhibit intersubjective validity.

Based on psychophysical and electrophysiological considerations, Blauert and Cobben [8] proposed a running cross correlation (RCC) of recorded audio signals

[15] See Mason et al. [36].

$$\mathrm{RCC}(\mathbf{r}, t, \tau) = \int_{-\infty}^{t} q_\mathrm{L}(\mathbf{r}, \delta) q_\mathrm{R}(\mathbf{r}, \delta + \tau) G(\mathbf{r}, t - \delta) \mathrm{d}\delta \, . \tag{10}$$

Here, q is the recorded signal p after applying a half-wave rectification and a smoothing in terms of low-pass filtering. The RCC is a function of time and lag, so it yields one cross correlation function for each time step. $G(\mathbf{r}, t - \delta)$ is a weighting function to attenuate past values

$$G(s) = \begin{cases} e^{\frac{-s}{5\,\mathrm{ms}}} & \text{for } \begin{cases} s \geq 0 \\ s < 0 \end{cases} \\ 0 \end{cases} . \tag{11}$$

The RCC produces peaks that are in fair agreement with lateralization judgments and the precedence effect, i.e. a dominance of the first wavefront. But the authors emphasize the need for improvements.

Yanagawa and Tohyama [47] conducted an experiment with a leading sound and a delayed copy of it, simulating direct sound and one reflection. They found that the interaural correlation coefficient (ICC) is a better estimator of source broadening than BQI. The ICC equals the ICCF, Eq. 2, when τ is chosen to be 0. Lindemann [33] uses the same measure but divides the signal into several frequency bands. He hypothesizes that small differences between the perceived location of frequency bands are the reason for subjective diffuseness.

Blauert and Lindemann [9] found evidence that early reflections with components above 3 kHz create an image expansion. But Bradley et al. [10] have found that late arriving reflections may again diminish ASW. However, the idea of ASW is that a listener is rather far away from the source. Consequently, the original width of a musical instrument is in the order of one degree or less. This original sound source is "extended" due to a decorrelation of ear signals which are caused by unsymmetrical reflections. But when being close enough to a musical instrument, it does have a notable width of many degrees. This width can be heard. In proximity to a source, direct sound already creates decorrelated signals at both ears. This decorrelation mainly results from the frequency- and direction-dependent radiation characteristics of musical instruments. Decorrelation of stereo and surround channels is common practice in music production to achieve the sensation of a broad sound source. In ambisonics and wave field synthesis, complex source radiation patterns are synthesized to create this impression. Source width in music production is discussed in the following section.

3 Source Width in Music Production

Perceived source width is of special interest in music production. In text books for recording, mixing and mastering engineers, spaciousness plays a major role. In the rather practical book written by Levinit [32], a chapter about recording tips and

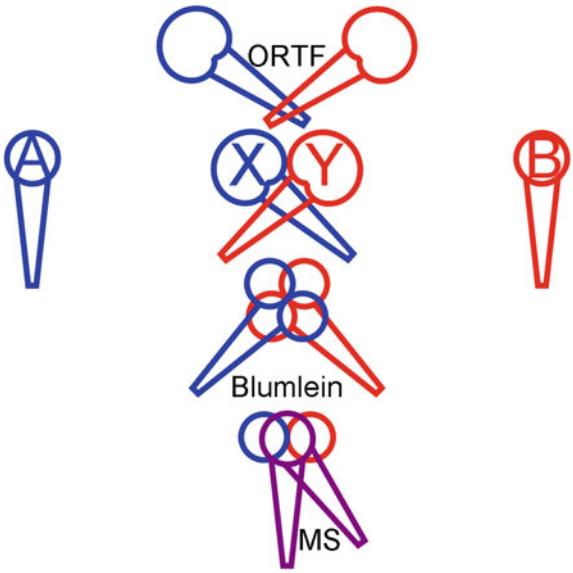


Fig. 1 Common stereo recording techniques

tricks has a section named "Making Instruments Sound Huge". Likewise, the audio engineer Kaiser [28] points out that the main focus in mastering lies in the stereo width, together with other aspects, such as loudness, dynamics, spaciousness and sound color.[16]

Probably by hearing experience, rather than due to fundamental knowledge of psychoacoustics and subjective room acoustics, sound engineers have found several ways to capture the width of musical instruments via recording techniques or to make them sound larger by pseudo-stereo methods. These are discussed in this section, followed by methods of source broadening in ambisonics and wave field synthesis application.

3.1 Source Width in Stereo and Surround

For recorded music, several microphoning techniques have been established. In the far field, they are used to capture the position of instruments in an ensemble and to record different portions of reverberation. In the near field, they capture the width of a solo instrument to a certain degree. Figure 1 shows some common stereo microphone techniques, namely A-B, Blumlein, mid-side stereo (MS), ORTF and X-Y. They are all based on a pair of microphones. The directivity of the microphones is depicted here by the shape of the head: omnidirectional, figure-of-eight

[16]See Kaiser [28], e.g. p. 23 and p. 40.

and cardioid. The color codes to what stereo channel the signal is rooted. Blue means left channel, red means right channel and violet denotes that the signal is routed to botch channels. Directional microphones that are placed closely together but point at different angles create mainly inter-channel level differences (ICLDs). This is the principle of X-Y recording. In A-B-recording, a large distance between microphones creates additional inter-channel time differences (ICTDs). So the recording techniques create systematically decorrelated stereo signals. The Blumlein recording technique creates even stronger ICLDs for frontal sources but more ambient sound or rear sources are recorded as well. In MS, sound from the neutral axis of the figure-of-eight microphone is only recorded by the omnidirectional microphones. It is routed to both stereo channels. The recording from the figure-of-eight microphone mainly captures lateral sound incidence and is added to the left and subtracted from the right channel. MS recording is quite flexible because the amplitude ratio between the monaural omnidirectional (mid-component) and the binaural figure-of-eight recording (side-component) can be freely adjusted. In all recording techniques, the degree of ICLD and ICTD depends on the position and radiation patterns of the source as well as on the amount and characteristics of the recording room reflections. More details on the recording techniques are given e.g. in Kaiser [27] and Friedrich [17].[17] It is also common to pick up the sound of musical instruments at different positions in the near field, for example with one microphone near the neck and one near the sound hole of a guitar. This is supposed to make the listener feel like being confronted with an instrument that is as large as the loudspeaker basis or like having the head inside the guitar.[18] When a recording sounds very narrow, it can be played by a loudspeaker in a reverberation chamber and recorded with stereo microphone techniques.[19] This can make the sound broader and more enveloping.

Recording the same instruments twice typically yields a stronger and, more importantly, dynamic decorrelation. Slight differences in tuning, timing, articulation and playing technique between the recordings occur. As a consequence, the relation of amplitudes and phases, transients and spectra changes continuously. These recordings are hard-panned to different channels, typically with a delay between them.[20] This overdubbing technique occurred in the 1960s.[21] Virtual overdubbing can be performed if the recording engineer has only one recording.[22] Adding one chorus effect to the left and a phase-inverted chorus to the right channel creates a dynamic decorrelation. In analog studios, artificial double tracking (ADT) was applied to create time-variant timing-, phase- and frequency differences between

[17]See especially Kaiser [27], pp. 33–43 and Friedrich [17], Chap. 13.

[18]This promise is made in Levinit [32], p. 157.

[19]See e.g. Faller [16].

[20]This is especially done for guitar and some vocal parts, see e.g. Kaiser [26], p. 116f and p.127 and Hamidovic [22], p. 57.

[21]See e.g. Maempel [34], p. 236.

[22]See e.g. Cabrera [11].

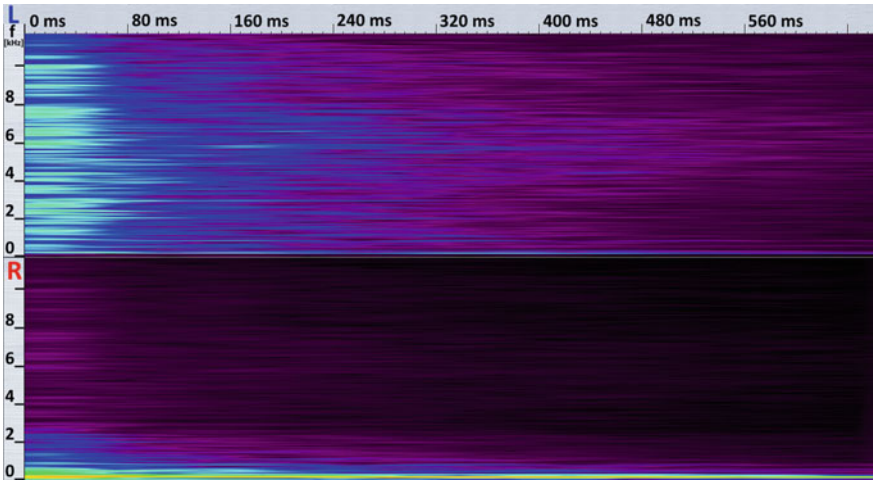


Fig. 2 Pseudostereo by high-passing the *left* and low-passing the *right* channel

channels. Here, a recording is re-recorded, using wow and flutter effects to alter the recording tape speed dynamically.

For electric and electronic instruments as well as for recorded music, several pseudostereo techniques are commonly applied to create the impression of a larger source. An overview of pseudo-stereophony techniques is given in Faller [16]. For example, sound engineers route a low-passed signal to the left and a high-passed signal to the right loudspeaker to increase the perceived source width as illustrated in Fig. 2. All-pass filters can be used to create inter-channel phase differences (ICPD) while maintaining a flat frequency response. Some authors report strong coloration effects, others less.[23] Usually, filters with a flat frequency and a random phase response are chosen by trial-and-error. Another method is to apply complementary comb filters[24] as indicated in Fig. 3. These create frequency-dependent ICLDs. Played back in a stereo setup, these ICLDs create ILDs but mostly to a lower degree, because both loudspeaker signals reach both ears. The ILDs are interpreted as different source angles by the listener. But, as long as the signals of the spatially spread frequency bands share enough properties, they remain fused. They are not heard as different source angles but as one spread source. Schroeder [43] investigated which sound parameters affect spatial sound impressions in headphone reproduction. He comes to the conclusion that ILD of spectral components have a greater effect on the perception of source width than IPD. Often, an ICTD between 50 and 150 ms is used to create a wide source. Sometimes, the delayed and attenuated copy of the direct sound is directly routed to the left channel and phase-inverted for the right. Applying individual filters or compressors for each

[23]See e.g. Cabrera [11] and Zotter and Frank [54] versus Faller [16].

[24]See e.g. Cabrera [11] and Kaiser [28], p. 154.

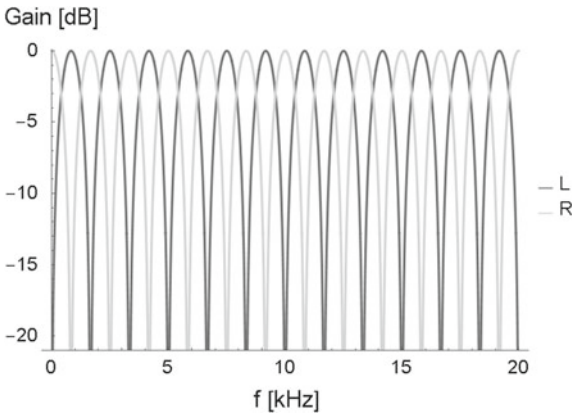


Fig. 3 Pseudostereo by applying complementary comb filters on the left and the right channel

channel is common practice, as well as creating a MS stereo signal and compressing or delaying only the side-component.[25] Likewise, it is very common to apply complementary equalizers to increase separation between instruments in the stereo panorama or to pan the reverb to a location other than the direct sound.[26] One additional way to create a higher spaciousness is to use a Dolby surround decoder on a stereo signal. This way, one additional *center channel* and one *rear channel* are created. These can be routed to different channels in a surround setup. The first is basically the sum of the left and the right channel whereas the latter is their difference, which is high-passed and delayed by 20–150 ms. This effect is called *magic surround*.[27] A general tip for a natural stereo width is to make bass frequencies most mono, mid-range frequencies more stereo and high frequencies most stereo,[28] i.e. with an increasing decorrelation of channels.

All of the named pseudo-stereo techniques are based on the decorrelation of loudspeaker signals. The idea is that the resulting interaural correlation is proportional to channel correlation. There are only few monaural methods to increase perceived source width. One practice is to simply use a compressor. The idea is inspired by the auditory system which, because of the level-dependent cochlear gain reduction, in fact operates as a 'biological compressor'. So a technical signal compressor creates the illusion that a source is very loud, and consequently very proximate to the listener. Naturally, proximate sources are wider, i.e. they are spread over more degrees from the listeners' point of view. Especially low frequencies should be compressed with a high attack time.[29]

[25]See Hamidovic [22], p. 57 and Kaiser [28], p. 152 and 156.

[26]See Kaiser [26], p. 50 and pp. 57f.

[27]See e.g. Faller [16] and Slavik and Weinzierl [45], p. 624.

[28]See Kaiser [28], pp. 148f.

[29]See e.g. Levinit [32], p. 158 and Rogers [41], p. 35.

Faller [16] proposes two additional pseudo-stereophony methods. The first is to compare a mono recording to a modern stereo mix and then create the same ICTD, ICLD and ICC for every subband. The second is to manually select auditory events in the spectrogram of the mono file and apply panning laws to spread instruments over the whole loudspeaker basis. Zotter and Frank [54] systematically alter inter-channel amplitude or phase differences of frequency components to increase stereo width. They found that the inter-channel cross correlation (ICCC) is approximately proportional to IACC in a range from $\mathrm{IACC}_u = 0.3$ to $\mathrm{IACC}_o = 0.8$. For both amplitude and phase alterations, they observe audible coloration.[30] Laitinen et al. [31] utilize the fact that in reverberant rooms, in contrast to anechoic conditions, the interaural coherence decreases with increasing distance to a sound source. This is not surprising as the direct-to-reverberant energy ratio (D/R ratio) decreases. The direct sound, which creates relatively high interaural coherence, is attenuated whereas the intensity of the relatively diffuse reverberance remains the same. Likewise, loudness and interaural phase coherence decreases with increasing distance to the source. They present formulas to control these three parameters. Gain factors are derived simply from listening to recreate the impression of three discrete distances. Control over perceived source distance might be related to perceived source extent.

In recording studios, a typical analyzing tool is the so-called *phase scope*, *vectorscope* or *goniometer*, plotting the values of the last x samples of the left versus the right channel as discontinuous Lissajous figures and additionally giving the inter-channel cross correlation coefficient.[31] This analysis tool is applied to monitor stereo width. It is illustrated in Fig. 4. The inter-channel cross correlation coefficient informs about mono compatibility. A negative correlation creates destructive interference when summing the stereo channel signals to one mono channel. When the left and right channel play the same signal, the goniometer shows a straight line. If amplitude differences occur, the line is deflected towards the channel with the louder signal. The more complicated the relation between the channel signals, the more chaotic the goniometer plot looks.

For surround systems with 5 or more channels, multi directional amplitude panning (MDAP) has been proposed. The primary goal of MDAP is to solve the problem of discontinuity: When applying amplitude based panning between pairs of loudspeakers, the perceived width of phantom sources is larger in the center and becomes more narrow for phantom source positions that are close to one of the loudspeakers. To increase the spread of lateral sources at least one additional speaker is activated. The principle is illustrated in Fig. 5. A target source width is chosen. It has to be at least the distance of two neighboring loudspeakers. One phantom source is panned to the left end of the chosen source extent, one phantom

[30]See Zotter and Frank [54] for details on their channel decorrelation methods and their investigations of IACC and sound coloration.

[31]See e.g. Kaiser [28], pp. 48ff although the meaning of the correlation coefficient is obviously misunderstood by this practician.

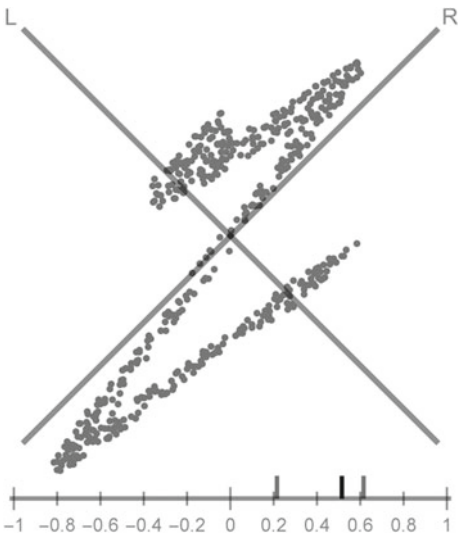


Fig. 4 Phase space diagram (*top*) and correlation coefficient (*bottom*) as objective measures of stereo width and mono compatibility

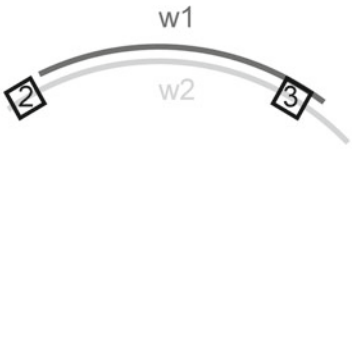


Fig. 5 Multi dimensional amplitude panning for different source widths

source is panned to the right end. For the illustrated source $w1$, loudspeakers 2, 3 and 4 are active. Source $w2$ has the same central source angle but a wider source extent. Here, loudspeaker 1 is additionally active.

3.2 Source Width in Ambisonics

Ambisonics started as microphone and playback technique in the 1970s. Pioneering work has been done by Gerzon.[32] The basic two-dimensional ambisonics recording technique is illustrated in Fig. 6. It is referred to as *first order ambisonics*.

[32]See e.g. Gerzon [19].

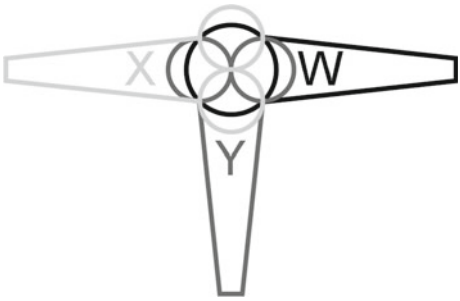


Fig. 6 First order ambisonics recording technique

One pressure microphone W and two perpendicular pressure gradient microphones X and Y are used. In the three-dimensional case, an additional figure-of-eight microphone captures the pressure gradient along the remaining axis, referred to as *B-Format* or *W, X, Y, Z*. Three-dimensional audio is out of scope of this chapter.

In contrast to conventional stereo recording techniques, the signals are not directly routed to discrete loudspeakers. They rather encode spatial information, namely the pressure distribution on a circle. The three microphones perform a truncated circular harmonic decomposition of the sound field at the microphone position. The monopole recording W gives the sound pressure at the central listening position p_0, i.e. the circular harmonic of 0th order. It is routed directly to the zeroth channel, i.e.

$$\text{ch0} = \frac{W}{\sqrt{2}} \,. \tag{12}$$

Recordings X and Y are the pressure gradients along the two spatial axes, i.e. 1st order circular harmonics. They can be approximated by

$$\text{ch1} = X \approx p_c(0) - p_c(\pi) \tag{13}$$

and

$$\text{ch2} = Y \approx p_c\left(\frac{\pi}{2}\right) - p_c\left(\frac{3\pi}{2}\right) \,. \tag{14}$$

Here, $p_c(\phi)$ are omnidirectional recordings of microphones that are distributed along a circle with a small diameter. Higher order encoding can be performed with more pressure receivers. For an encoding of order n, $4n + 1$ pressure receivers are necessary. Figure 7 illustrates ambisonics recordings of different orders for the same wave field. Recordings from microphones at different angles are combined like

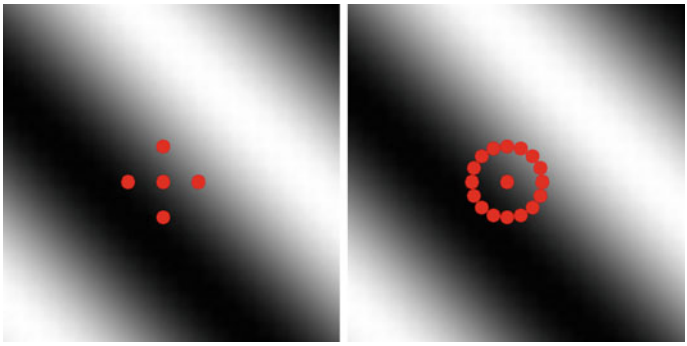


Fig. 7 1st order (*left*) and 4th order (*right*) ambisonics recording of a plane wave

$$\text{ch3} \approx p_c(0) - p_c\left(\frac{\pi}{2}\right) + p_c(\pi) - p_c\left(\frac{3\pi}{2}\right) \tag{15}$$

and

$$\text{ch4} \approx p_c\left(\frac{\pi}{4}\right) - p_c\left(\frac{3\pi}{4}\right) + p_c\left(\frac{5\pi}{4}\right) - p_c\left(\frac{7\pi}{4}\right) . \tag{16}$$

Figure 8 illustrates the circular harmonics. Their superposition yields the *n*th order approximation of the sound field along the circle. The first order approximation yields a cardioid. The maximum points at the incidence angle of the wave front. The lobe is rather wide. In contrast to that, the maximum of the 4th order approximation is a relatively narrow lobe that points at the incidence angle of the wave front. However, several sidelobes occur. The order gives the precision with which the sound field is encoded. For one plane wave, the first order approximation already yields the source angle. For superimposed sound fields with several sources and complicated radiation patterns, a higher order is necessary to encode the sound field adequately. However, a finite order might always contain artifacts due to sidelobes.

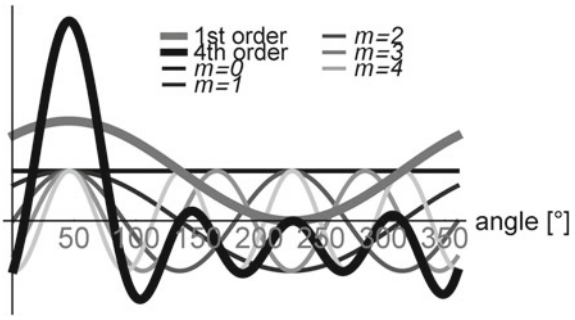


Fig. 8 Circular harmonics of order 0 and 1 are encoded in 1st order ambisonics. In 4th order ambisonics, additional circular harmonics of order 2, 3 and 4 are necessary

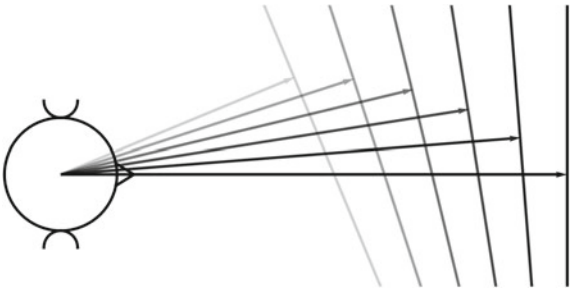


Fig. 9 Phantom source widening in ambisonics by synthesizing frequency dispersed source positions. Different frequency regions are indicated by different gray levels

Ambisonics decoders use different strategies to synthesize the encoded sound field at the central listening position. This is either achieved by the use of projection or by solving a linear equation system that describes the relationship between loudspeaker position, wave propagation and the encoded sound field on a small circle around the central listening position. Ambisonics decoders are out of scope of this chapter. An overview can be found e.g. in Heller [23].

Zotter et al. [55] propose a method which is related to the idea of a frequency-dependent MDAP. In an ambisonics system, frequency regions are not placed at the same source position but spread over discrete angles. In a way, this is a direct implementation of the hypothesis that has been formulated by Lindemann [33] who believes that deviant source localizations of different frequency bands is the reason for subjective diffuseness. The principle is illustrated in Fig. 9. In their listening test, the perceived source extent, reported by 12 subjects, correlated with the BQI when increasing the time lag to $\tau = 2$ ms.[33]

Another principle is tested in Potard and Burnett [40]. They synthesize 6 virtual point sources with 4th order ambisonics. The virtual source positions are spread over different angles. White noise is divided into three frequency bands. The signal for each virtual point source is composed of decorrelated versions of these frequency bands. The decorrelation is achieved by all pass filters. Then, they mix each frequency band of the original source signal with the decorrelated version. With the mixing ratio ξ and the distribution of the virtual point sources, they try to control the source width of each frequency region. The perceived source extents reported by 15 subjects are in fair agreement with the intended source extents. Unfortunately, no systematic alteration of virtual source spread and degrees of decorrelation are presented in their work.

The authors in Laitinen et al. [30] propose an implementation of directional audio coding (DirAC) in ambisonics. A premise of their approach is that the human auditory system perceives exactly one direction and one source extent for each frequency band in each time frame. From an ambisonics recording they derive the source angle and its diffuseness in terms of short-term fluctuations or uncertainty.

[33]Their approach and experiment are documented in Zotter et al. [55]. The information that thetime lag was increased cannot be found in the paper; it was given verbally at the conference.

The source angle is created by ambisonics decoding. Diffuseness is created by decorrelated versions that are reproduced by different loudspeakers. In a listening test with 10 subjects, they found that localization and sound quality were very good with their approach. For future research, they propose to investigate the perceived source extent in more detail.

Just as in stereo, the presented ambisonics approaches either aim at controlling the signals at discrete channels or at controlling the spatial spread of virtual sources. Focusing on the sound field at the listening position might reveal a deeper insight into the relationship between ear signals and the perception of width. This is not the case for all wave field synthesis techniques. These are discussed in the following.

3.3 Source Width in Wave Field Synthesis

Wave field synthesis is based on the idea that the sound field within an enclosed space can be controlled by signals on its surface. An overview of its theory and application can be found in Ziemer [51]. Typically, wave fronts of static or moving virtual monopole sources or plane waves are synthesized in an extended listening area. With this procedure, listeners experience a very precise source location which stays stable, even when moving through the listening area. However, due to the simple omnidirectional radiation pattern, virtual sources tend to sound small. This observation called several researchers into action, trying to make sources sound larger, if desired.

Baalman [3][34] arranged a number of virtual point sources to form a sphere, a tetrahedron and an icosahedron, each with a diameter of up to 3.4 m. With this distribution of virtual monopole sources, she played speech and music to subjects. The shapes were perceived as being further away and broader than a monopole source. The most perceivable difference was the change in tone color. In her approach the perceived source width did not depend on the width of the distributed point sources. There are several potential reasons why her method failed to gain control over perceived source widths. One reason might be that the distributed point sources radiated the same source signal. No filtering or decorrelation was performed. Except for low frequencies, coherent sound radiation from all parts of a source body is rather unusual and does not create the perception of a large source width. Wave field synthesis works with exactly this principle; delayed and attenuated versions of the same source signal are played by a closely spaced array of loudspeakers to recreate the wave front of a virtual monopole source or plane wave. Thus, the difference between one virtual monopole and a spherical distribution of coherent virtual monopoles can only lie in synthesis errors and in comb filter effects that depend on the distance of the point sources. Another reason might have been that the distance between listeners and source was in all cases more than 3 m. So

[34]See Baalman [3], Chap. 7.

Fig. 10 Combined (*left*) and plain (*right*) multipoles of low orders

when measuring source width in degrees, the shapes are again relatively narrow in most trials.

In Corteel [12], the synthesized sources are no monopoles but circular harmonics of order 1–4 and some combinations of those, i.e. multipoles. Some exemplary radiation patterns are illustrated in Fig. 10. The paper focuses on the optimization of filters to minimize physical synthesis errors. It does not include listening tests that inform about perceived source extent. However, as soon as a multipole of low order is placed further than a few meters away from a listener, it barely creates interaural sound differences. The reason is that multipoles of low order are very smooth. Assuming a distance of 0.15 m between the ears, the angle between the ears and a complexly radiating point source at 3 m distance is about 2.8°. Only slight amplitude and phase changes occur over this angle width for low order multipoles. This can easily be seen in Fig. 10. For steep, sudden changes to occur within a few degrees, a very high order is necessary.

In Jacques et al. [25], single musical instruments or ensembles are recorded with a circular microphone array consisting of 15 microphones. They synthesize the recordings by means of virtual high order cardioid sources, pointing away from the origin, i.e. the original source point. This way, the radiation pattern is reconstructed to a certain degree. In a listening test, subjects were able to hear the orientation of a trumpet with this method. When synthesizing only one high order cardioid, many subjects had troubles localizing the source. This was, however, not the case when several high order cardioids reconstruct an instrument radiation pattern.

In Ziemer and Bader [53], the radiation characteristic of a violin is recorded with a circular microphone array which contains one microphone every 2.8°. The radiation characteristic is synthesized in a wave field synthesis system. This is achieved by simplifying the violin as complex point source. The physical approach is the same as in the present study and will be explained in detail in Sect. 4.2. The main aim of this paper is to utilize psychoacoustic phenomena to allow for physical synthesis errors while ensuring precise source localization and a spatial sound impression. In a listening test with 24 subjects, the recreated violin pattern could be localized better than a stereo phantom source with plain amplitude panning. Still, it was perceived as sounding more spatial.

The approach to model virtual sources with more complex radiation characteristics to achieve control over ASW is very promising. But it is necessary to create the cues that affect ASW. These cues are to be created by the virtual source and by synthesized reflections. But more important than the sound field at the virtual source position is the sound field at the ears of the listener. In the study that is

described in the following section, relationships between source width and the sound field at listening positions are investigated.

4 Sound Radiation and Source Extent

In this investigation the actual extent of the vibrating part of certain musical instruments is related to quantities of the radiated sound. Here, the focus lies on direct sound. The idea behind this procedure is straightforward: There must be evaluable quantities in the radiated sound that indicate source width because the auditory system has no other cues than these. As mentioned earlier, investigations which aimed at explaining perceived source width of direct sound by controlling signals of loudspeakers—instead of the signals at listeners' ears—did not succeed. But if we find parameters in the radiated sound that correlate with actual physical width we may have found the cues which the auditory system consults to render a judgment about source width. By controlling these parameters, more targeted listening tests can be conducted. Furthermore, when the relationship between audio signal and width perception is disclosed, it can be implemented as a tool for stereo, ambisonics and wave field synthesis applications to control perceived source extent.

This investigation is structured as follows: First, the setup to measure the radiation patterns of musical instruments is introduced and the examined instruments are listed. Then, the complex point source model is briefly described. The model is applied to propagate the instrumental sound to several potential listening positions. For these listening positions, physical sound field quantities are calculated. Basically, the quantities are taken from the field of psychoacoustics and subjective room acoustics. But they are adopted to free field conditions and instrumental sounds. The adopted versions are discussed subsequently. Finally, relationships between sound field quantities and the physical source extent are shown. It is demonstrated how a combination of two parameters can be used to predict the source extent. Although physical sound field quantities are put into relation with physical source extent, the findings allow some statements about psychoacoustics. So the results are discussed against the background of auditory perception. Potential applications and future investigations are proposed in the prospects section.

4.1 Measurement Setup

In an anechoic chamber a circular microphone array was installed roughly in the height of the investigated musical instruments. It contains 128 synchronized electret microphones. An instrumentalist is placed in the center, playing a plain low note without strong articulations or modulations, like vibrato or tremolo. One second of quasi-stationary sound was transformed into the spectral domain by discrete Fourier transform (DFT) yielding 128 complex spectra

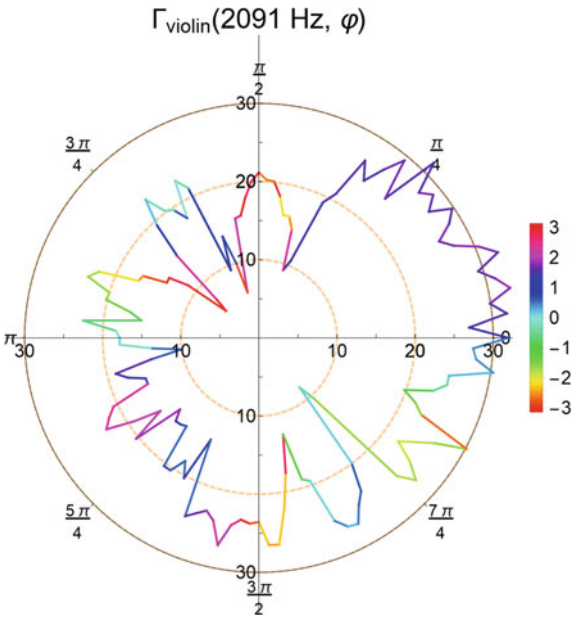


Fig. 11 Measured radiation pattern of one violin frequency

$$P(\boldsymbol{\omega}, \mathbf{r}) = \mathrm{DFT}[p(t, \mathbf{r})] \tag{17}$$

where $\mathbf{r}$ is the position vector of each microphone, consisting of its distance to the origin $\mathbf{r}$ and the angle ϕ between the microphone and the normal vector which is the facing direction of the instrumentalist. Each frequency bin in a complex spectrum has the form $\hat{A}e^{i\varphi}$ with the amplitude $\hat{A}$, the phase φ, Euler's number e and the imaginary unit i. The complex spectra of one violin partial are illustrated in Fig. 11. The amplitude is plotted over the corresponding angle of the microphones, the phase is coded by color. With this setup the radiated sound of 10 instruments has been measured. The investigated instruments are listed in Table 1. Just as in most room acoustical investigations, only partials up to the upper limit of the 8 kHz octave band, i.e. $f_{\max} = 11,314\,\mathrm{kHz}$, are considered. For higher frequencies, the density of partials becomes very high and the signal-to-noise ratio becomes low. Partials are selected manually from the spectrum to find partials, double peaks and to exclude electrical hum etc. reliably.

4.2 The Complex Point Source Model

To compare these musical instruments despite their mostly dissimilar geometries, they are simplified as complex point sources for further investigations. In principle, the complex point source model can be explained easily by Figs. 12 and 13.

Table 1 List of investigated musical instruments and their width at three different distances

Instrument	Width (°)
Accordion	28/19/10
Bagpipe	23/15/8
Crash cymbal	37/25/13
Dizi flute	11/8/4
Double bass	36/24/12
Harmonica	13/9/4
Mandolin	35/24/12
Shakuhachi	11/8/4
Tenor saxophone	11/8/4
Violin	19/13/6

The crash cymbal and the dizi flute have been added after the presentation of preliminary results in Ziemer [50]

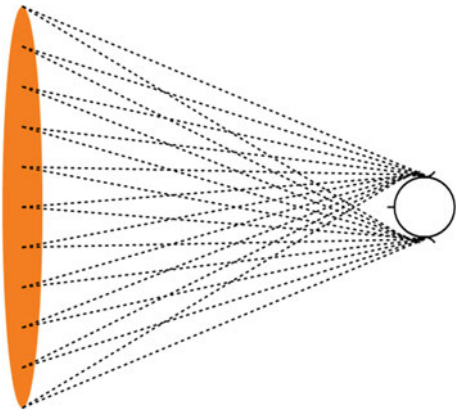


Fig. 12 Schematic sound path from an extended source to the ears. The superposition of radiated sound from all parts of the instrumental body reach both ears

Figure 12 shows a sampled version of the paths that pressure fluctuations undergo from the surface or enclosed air of an extended source to the ears of a listener. Radiations from all parts of the instrument reach both ears. In this consideration we neglect near field effects like evanescent waves and acoustic short circuits. Figure 13 shows a drastic simplification. The instrument is now considered as one point which radiated sound towards all direction, modified by the amplitude and phase that we have measured for the 128 specific angles.

The radial propagation of a point source can be described by the free field Green's function

$$G(r) = \frac{\mathrm{e}^{-ikr}}{r}, \tag{18}$$

where the pressure amplitude decays according to the $1/r$ distance law and the phase shifts according to the wave number $k = 2\pi/\lambda$, where λ is the wave length.

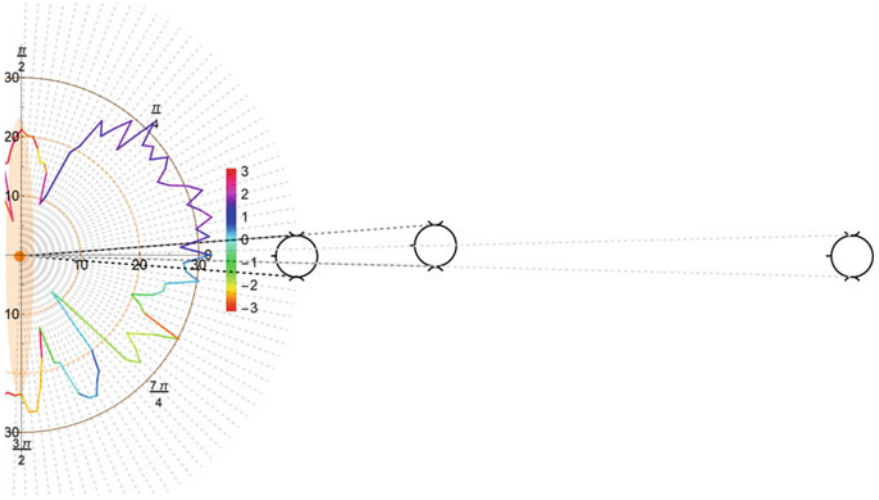


Fig. 13 Ear signals resulting from the complex point source simplification

Covering a circumference with 128 microphones yields one microphone every $\Delta\phi = 2.8°$. The distance between the two ears of a human listener is about 0.15 m. Assuming a listener facing the source point at a distance of 1 m, the distance of the ears correspond to every third microphone, at a distance of 1.5 m every second microphone and at 3 m every microphone. Thus, we can calculate interaural signal differences by comparing every third recording or by propagating all measured signals to a distance of 1.5 and 3 m by Eq. 18 and compare every second or every neighboring propagated microphone recording. This yields a set of $3 \times 128 = 384$ virtual listening positions for which we can calculate ear signals without the use of interpolations.

Neglecting the actual source geometry and considering a musical instrument as a point instead is a rather drastic simplification. Still, the computational benefits are obvious. Furthermore, the model has proven to yield plausible results both physically and perceptually.[35]

4.3 Physical Measures

For all 384 virtual listening positions a number of monaural and binaural physical measures has been calculated. Although no actual listeners are present, the measured and propagated microphone signals are termed "ear signals" in this investigation. Most of them are derived from parameters used in the field of

[35]As has been reported e.g. in Ziemer [49], Ziemer and Bader [52] and Otondo and Rindel [38].

psychoacoustics or room acoustics. But they are adapted to pure, direct, instrumental sound. Due to the vast consensus in the literature,[36] a combination of one monaural and one binaural parameter is searched which best predict the width of musical instruments. The monaural parameter quantifies the strength of bass, the binaural parameter represents the portion of interaural differences compared to interaural coherence. Monaural and binaural parameters are described subsequently.

4.3.1 Monaural Measures

The early low strength $G_{E,\mathrm{low}}$—mentioned in Sect. 2.2, Eq. 5—cannot be applied to pure direct sound as it is the ratio of bass energy in the reverberant field compared to the free field. Therefore, other parameters have been tested, representing the relative strength of low frequencies.

First, all partials f_i below $f_{\max} = 11.314\,\mathrm{kHz}$ are selected manually from the spectrum. As a monaural measure, the fundamental frequency f_1 of each instrumental sound is determined. Likewise, the number of partials I present in the considered frequency region is counted. For harmonic spectra that contain all multiple integers of the fundamental, I should be proportional to $1/f_1$. This is not the case for inharmonic spectra like that of the crash cymbal or instruments like the accordion, which show beatings, i.e. double peaks. Thus, both measures are considered as potential monaural descriptors for a multiple regression analysis. These quantities characterize the source spectrum. They are independent of the listening position.

The amplitude ratio between partials in the 125 and 250 Hz octave bands and in the 500 and 1000 Hz octave bands quantifies bass as a *bass ratio* (BR). A linear and a logarithmic bass ratio

$$\mathrm{BR}_{\mathrm{lin}}(\phi) = \frac{\sum_{f_i \geq 88\,\mathrm{Hz}}^{f_i < 355\,\mathrm{Hz}} \hat{A}^2(f_i)}{\sum_{f_i \geq 355\,\mathrm{Hz}}^{f_i \leq f_{\max}} \hat{A}^2(f_i)} \tag{19}$$

and

$$\mathrm{BR}_{\mathrm{log}}(\phi) = \frac{\sum_{f_i \geq 88\,\mathrm{Hz}}^{f_i < 355\,\mathrm{Hz}} 10\lg\left(\frac{\hat{A}^2(f_i)}{\hat{A}^2(f)_{\min}}\right)}{\sum_{f_i \geq 355\,\mathrm{Hz}}^{f_i \leq f_{\max}} 10\lg\left(\frac{\hat{A}^2(f_i)}{\hat{A}^2(f)_{\min}}\right)} \tag{20}$$

are calculated. Here, $\hat{A}^2(f)_{\min}$ is the lowest amplitude of all partials found in the four octave bands. These two parameters are similar to the *bass ratio* known from room acoustics. In room acoustics, typically reverberation times, early decay times or, sometimes, strength of low frequencies are compared to midrange frequencies.

[36] Refer to the literature cited in Sect. 2.2.

As some instruments create even lower frequencies, and most instruments create much higher frequencies, these two measures can be extended to a relative *bass pressure* (BP) and *bass energy* (BE) in the sound:

$$\mathrm{BP}(\phi) = \frac{\sum_{i=1}^{f_i < 355\,\mathrm{Hz}} \hat{A}(f_i)}{\sum_{f_i \geq 355\,\mathrm{Hz}}^{f_i \leq f_{\max}} \hat{A}(f_i)} \tag{21}$$

$$\mathrm{BE}(\phi) = \frac{\sum_{i=1}^{f_i < 355\,\mathrm{Hz}} \hat{A}^2(f_i)}{\sum_{f_i \geq 355\,\mathrm{Hz}}^{f_i \leq f_{\max}} \hat{A}^2(f_i)} \tag{22}$$

For BP the sum of amplitudes $\hat{A}(f_i)$ of all frequencies below the upper limit of the 250 Hz octave band is compared to the sum of all other considered partials' amplitudes. This value is similar to *BE*, which is the ratio of squared amplitudes. Note that BP^2 does not equal *BE*. If only low-frequency sound is present, all four ratios are undefined as the denominator would be zero. In all other cases they are positive values. The higher the value the higher the sound pressure of the low-frequency components compared to higher partials.

The functions of BE and $\mathrm{BR}_{\mathrm{lin}}$, plotted over the angle, look quite similar. An example is shown in Fig. 14. Especially when transforming the values to a logarithmic scale, BE, $\mathrm{BR}_{\mathrm{lin}}$ and BP look rather similar. This can be seen in Fig. 15, where the logarithm of the three quantities is plotted over angle and scaled to similar magnitudes.

As the monaural parameter is supposed to represent the presence or strength of bass, the *spectral centroid* is a meaningful measure. According to Shimokura et al. [44], *C* is strongly related to the spectral distribution and to W_{IACC}, which had been

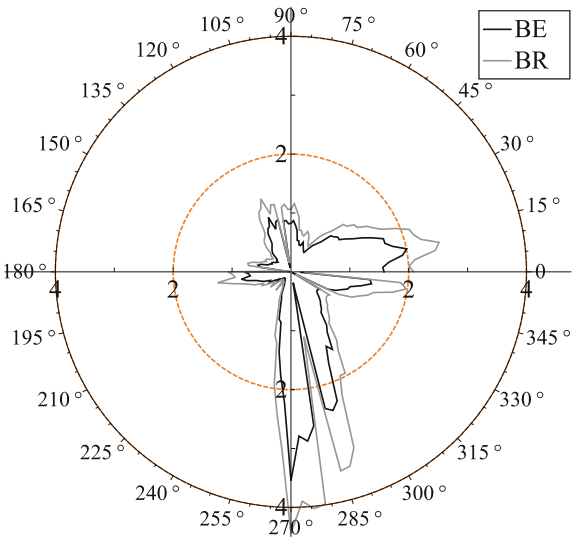


Fig. 14 BE and $\mathrm{BR}_{\mathrm{lin}}$ of a bagpipe, plotted over the listening angle

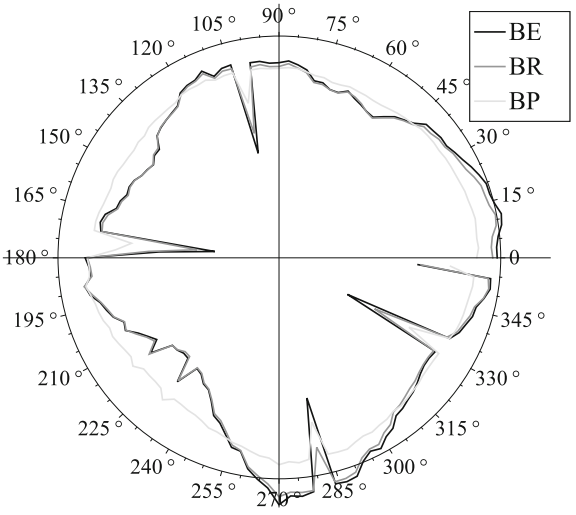


Fig. 15 Logarithmic plot of BE, BR_{lin} and BP of a bagpipe. They are scaled to similar magnitudes

proposed to quantify bass in ASW investigations. Three versions of the spectral centroid are calculated, namely the classic spectral centroid

$$C(\phi) = \frac{\sum_{f=20\,\mathrm{Hz}}^{20\,\mathrm{kHz}} f\hat{A}(f,\phi)}{\sum_{f=20\,\mathrm{Hz}}^{20\,\mathrm{kHz}} \hat{A}(f,\phi)}, \tag{23}$$

where all spectral components are included. The upside of this measure is that even higher partials and noisy components are considered. The downside is that this measure is sensitive to noise of the measurement equipment. This sensitivity is reduced when limiting the bandwidth to the octave bands from 63 Hz to 8 kHz, to get the *band-passed spectral centroid*

$$C_{\mathrm{bp}}(\phi) = \frac{\sum_{f=43\,\mathrm{Hz}}^{11{,}314\,\mathrm{Hz}} f\hat{A}(f,\phi)}{\sum_{f=43\,\mathrm{Hz}}^{11{,}314\,\mathrm{Hz}} \hat{A}(f,\phi)}. \tag{24}$$

The most robust approach is to calculate the spectral centroid only from all manually selected partials

$$C_{\mathrm{part}}(\phi) = \frac{\sum_{i=1}^{I} f_i\hat{A}(f_i,\phi)}{\sum_{i=1}^{I} \hat{A}(f_i,\phi)}. \tag{25}$$

These monaural quantities are independent of the listening distance but they depend on listening angle. Therefore, the mean value over all angles is taken.

In summary, the nine monaural parameters f_1, I, BR_{lin}, BR_{lin}, BP, BE, C, C_{bp} and C_{part} are determined. Monaural measures are independent of the listening

distance whereas source width in degrees is not. Hence, no high correlation between monaural parameters and source extent is expected.

4.3.2 Interaural Measures

As stated before, interaural signal differences are expected to have a larger contribution to width perception than monaural cues. They are calculated from the signals that have been recorded at or propagated to the ear positions of the 384 virtual listeners.

Following the idea of the lateral energy fraction (LEF_{E4}), Eq. 1, the binaural pressure component (BPC) is proposed as the mean ratio between interaural and monaural sound pressure component of all partials

$$\mathrm{BPC}(\mathbf{r}) = \sum_{f_i \geq 88\,\mathrm{Hz}}^{f_i \leq 1{,}414\,\mathrm{Hz}} \frac{|P(f_i, \mathbf{r}_\mathrm{L}) - P(f_i, \mathbf{r}_\mathrm{R})|}{|P(f_i, \mathbf{r}_\mathrm{L}) + P(f_i, \mathbf{r}_\mathrm{R})|} / \mathrm{norm}. \tag{26}$$

for the octave bands from 125 to 1000 Hz. The norm is the bandwidth, i.e. the distance between the actual lowest and highest partial present within these four octave bands. Similarly, the binaural energy component (BEC)

$$\mathrm{BEC}(\mathbf{r}) = \sum_{f_i \geq 88\,\mathrm{Hz}}^{f_i \leq 1{,}414\,\mathrm{Hz}} \frac{(P(f_i, \mathbf{r}_\mathrm{L}) - P(f_i, \mathbf{r}_\mathrm{R}))^2}{(P(f_i, \mathbf{r}_\mathrm{L}) + P(f_i, \mathbf{r}_\mathrm{R}))^2} / \mathrm{norm}. \tag{27}$$

is the ratio between the squared sound pressure difference and the squared sum.

BPC and BEC of a dizi flute are plotted for all listening positions in Figs. 16 and 17. The BPC has higher values, in the BEC some peaks are emphasized compared to the BPC.

It is not meaningful to apply the binaural quality index (BQI), Eq. 4, to the direct instrumental sounds. In room acoustical investigations, the time lag accounts for the fact that lateral reflections might arrive at a listener. These create a maximum interaural time difference of almost ±1 ms. The time lag compensated for this interaural time difference. But under the present free field conditions, all virtual listeners face the source and no reflections occur. Thus, only the interaural correlation coefficient (ICC) is calculated. According to Yanagawa et al. [46], it is the better estimator of ASW, anyway. It equals Eq. 2 if τ is chosen to be 0.1—ICC of a mandolin is plotted in Fig. 18. The same fluctuations as in room acoustical investigations occur.

The interaural difference (IAD), Eq. 6, can be calculated for time windows of 40 ms just as proposed in Griesinger [20]. An example is plotted in Fig. 19. Like C, C_{bp}, and 1—ICC, this measure is sensitive to uncorrelated noise that is present in the recordings.

The ILD and IPD of one partial f_i can easily be calculated by

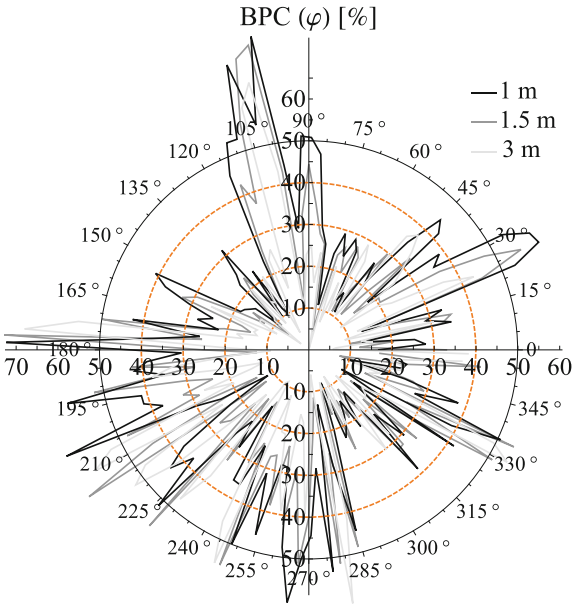


Fig. 16 Binaural pressure component (*BPC*) of a dizi flute at three listening distances plotted over listening angle

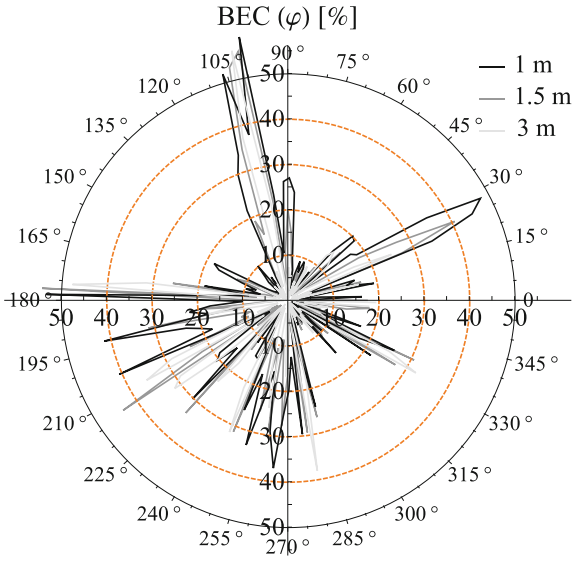


Fig. 17 Binaural energy component (*BEC*) of a dizi flute at three listening distances plotted over listening angle

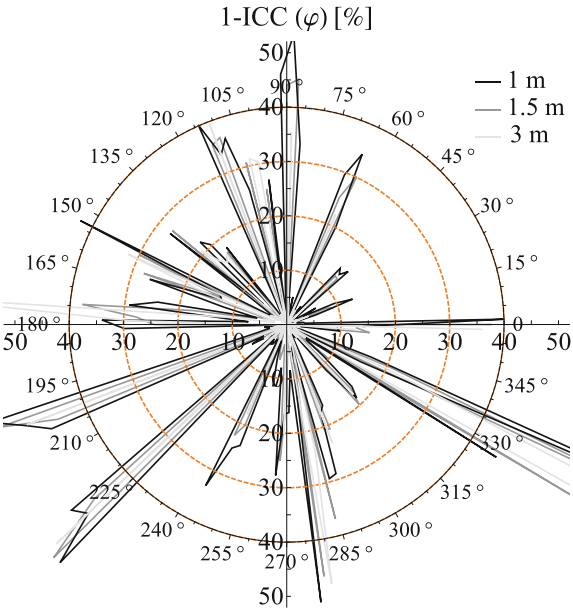


Fig. 18 1—ICC of a mandolin

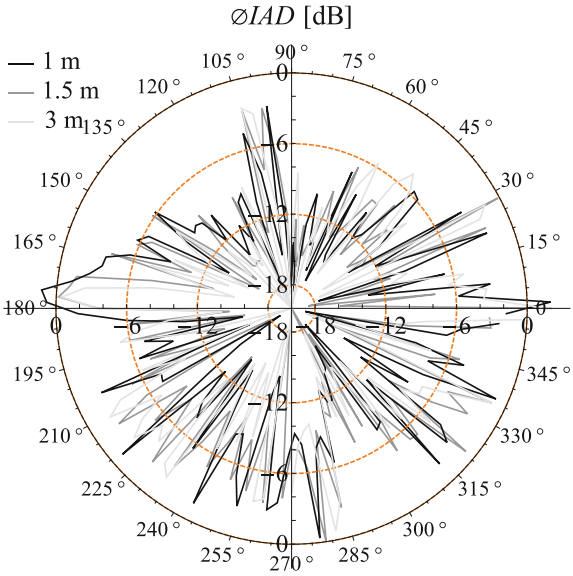


Fig. 19 IAD of a double bass

$$\mathrm{ILD}(f_i, \mathbf{r}) = \left| 20 \lg \left(\frac{\hat{A}(f_i, \mathbf{r}_L)}{\hat{A}(f_i, \mathbf{r}_R)} \right) \right| \tag{28}$$

and

$$\mathrm{IPD}(f_i, \mathbf{r}) = |\varphi(f_i, \mathbf{r}_L) - \varphi(f_i, \mathbf{r}_R)|. \tag{29}$$

Here, $\hat{A}$ is the amplitude and φ the phase. Naturally, the ILD and IPD of loud partials can be heard out more easily by a listener. Thus, they are expected to be more important than those of soft partials. Therefore, they are both weighted by the same factor

$$g(f_i, \mathbf{r}) = \frac{\left| \hat{A}(f_i, \mathbf{r}_L), \hat{A}(f_i, \mathbf{r}_R) \right|_\infty}{\hat{A}(\mathbf{r})_{\max}} \tag{30}$$

which is the larger amplitude of one frequency f_i at both ears L and R, normalized by the highest amplitude of all frequencies at the considered listening position $\hat{A}(\mathbf{r})_{\max}$. The factor g follows the idea of the binaural listening level LL which Ando [2] found to be important for width perception of multi-band noise. Combining Eq. 30 with 28 and 29, respectively, yields the weighted interaural level and phase difference (gILD and gIPD).

To be more close to human perception, the IPD parameter is adjusted by one more step. As mentioned above, the human auditory system is only sensitive to IPD below 1.2 kHz, so only partials below this upper threshold are considered to yield the weighted, band-passed interaural phase difference

$$g\mathrm{IPD}_{\mathrm{bp}}(f_i, \mathbf{r}) = g(f_i, \mathbf{r}) |\varphi(f_i, \mathbf{r}_L) - \varphi(f_i, \mathbf{r}_R)|, f_i \leq 1.2\,\mathrm{kHz}. \tag{31}$$

The evolution from IPD over gIPD to $g\mathrm{IPD}_{\mathrm{bp}}$ can be observed in Figs. 20, 21 and 22. These are plots of a harmonica. The IPD looks somewhat noisy and has two valleys around 20° and 200°. When weighting them with the amplitudes, gIPD looks quite similar. Only the overall magnitudes change. Neglecting all frequencies above 1.2 kHz, the magnitudes are even much lower. Some rather distinct peaks occur at several angles. These coincide with peaks in 1—ICC.

The main difference between the BQI and the $\mathrm{gIPD}_{\mathrm{bp}}$ lies in the fact that the former considers phase inversion not as spatial whereas the latter does. It is emphasized in Damaske and Ando [13] that if the maximum absolute value which determines the BQI comes from a negative value, the listening condition is unnatural.[37] This is evidence that ear signals being in phase and out of phase should be considered as being different in perception.

[37]See Damaske and Ando [13], p. 236.

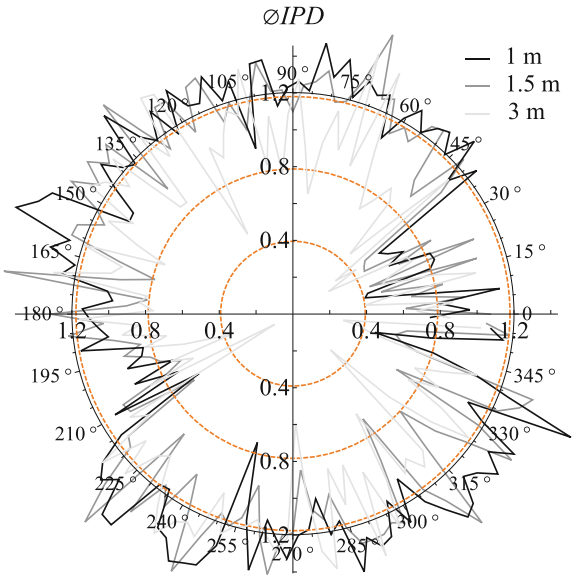


Fig. 20 IPD of the harmonica at all angles and distances

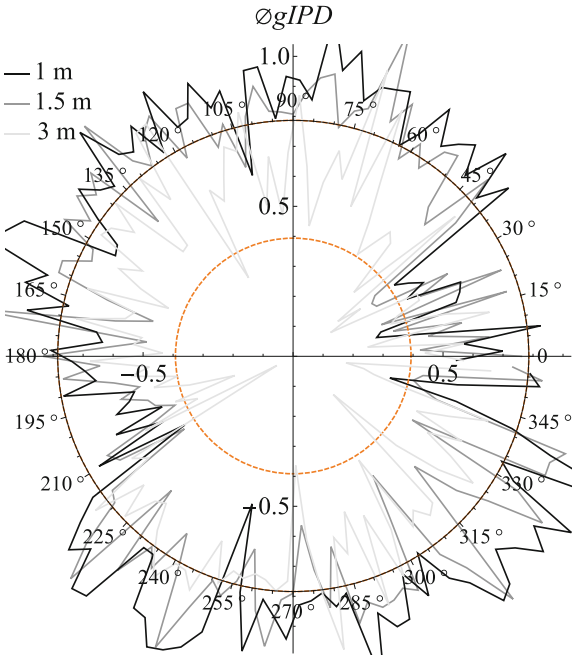


Fig. 21 gIPD of the harmonica at all angles and distances

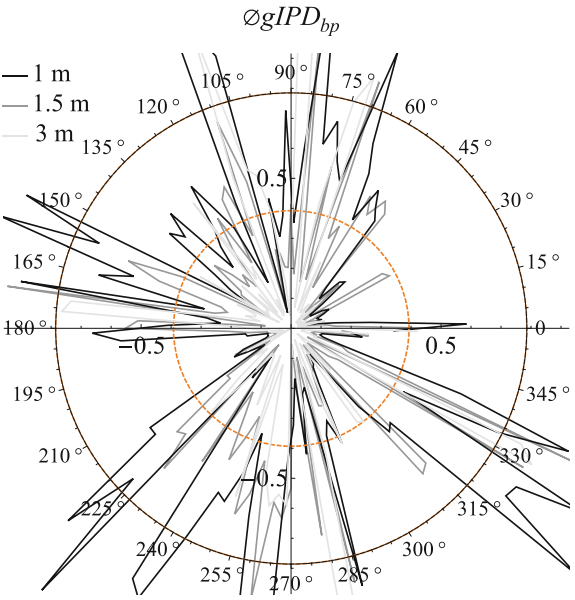


Fig. 22 $g\mathrm{IPD}_{\mathrm{bp}}$ of the harmonica at all angles and distances

In summary, the nine binaural sound field quantities BPC, BEC, 1—ICC, IAD, ILD, IPD, gILD, gIPD and $g\mathrm{IPD}_{\mathrm{bp}}$ are measured. As illustrated in the figures, these measures tend to have lower magnitudes at further distances. This is true for most angles. This behavior is expected, as the source width also decreases with increasing distance. Quantities like RL_{E}, Eq. 8, and $\mathrm{RCC}(t, \tau)$, Eq. 10, are not adopted to the present free field conditions. The first uses delay times of reflections, which are not present in this investigation. The latter assumes that the perceived source extent changes due to the amount and diffusion of reflections. This is not expected for a single note in a free field.

4.4 Results

All sound field quantities that exhibit a significant correlation with source width are listed in Table 2. Here, the Pearson correlation coefficient is listed. The significance level of $p < 0.05$ is indicated by bold numbers, $p < 0.01$ are underlined. Among the monaural measures, the lowest partial f_1, shows a significant negative correlation with width. The number of partials I in the considered frequency region exhibits a highly significant correlation with the source width ($p = 0.001830$). The scatter and the function of the linear regression are plotted in Fig. 23. The width is given in radian. One instrument creates three vertically arranged equidistant points. This is the case because it provides the same I for all three distances. The correlation between $\mathrm{BR}_{\mathrm{log}}$ and width lies slightly above the $p < 0.05$ level ($p = 0.060661$).

Table 2 Pearson correlation for all quantities that exhibit a significant correlation with width

	*f*1	*I*	BRlog	ILD	gILD	gIPDbp	1-CC	BPC	BEC	width
*f*1	1	**−0.676**	−0.135	−0.297	−0.083	0.017	**0.424**	−0.255	−0.342	**−0.396** 0.030350
I	**−0.676**	1	−0.224	**0.543**	0.356	0.051	0.019	**0.481**	**0.546**	**0.545** 0.001830
BRlog	−0.135	−0.224	1	0.048	**−0.409**	−0.208	**−0.704**	−0.047	−0.029	−0.365713 0.060661
ILD	−0.297	**0.543**	0.048	1	**0.620**	**0.442**	0.087	**0.586**	**0.483**	**0.449** 0.012866
gILD	−0.083	0.356	**−0.409**	**0.620**	1	**0.640**	**0.387**	**0.594**	**0.556**	**0.401** 0.028227
gIPDbp	0.017	0.051	−0.208	**0.442**	**0.640**	1	**0.383**	**0.807**	**0.725**	**0.591** 0.000581
1—ICC	**0.424**	0.019	**0.704**	0.087	**0.387**	**0.383**	1	0.248	0.147	**0.401** 0.028211
BPC	0.255	**0.481**	−0.047	**0.586**	**0.594**	**0.807**	0.248	1	**0.972**	**0.654** 0.000087
BEC	−0.342	**0.546**	0.029	**0.483**	**0.556**	**0.725**	0.147	**0.972**	1	**0.646** 0.000114

The significance levels $p < 0.05$ are bold, $p < 0.01$ are underlined. For width, the p-value is given below the correlation coefficient

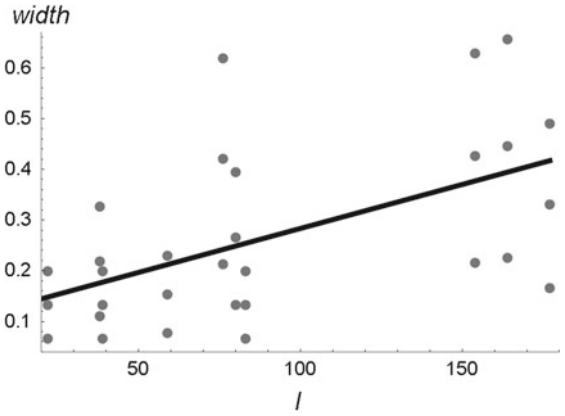


Fig. 23 Source width plotted over the number of partials I (*gray*) and the linear regression function (*black*)

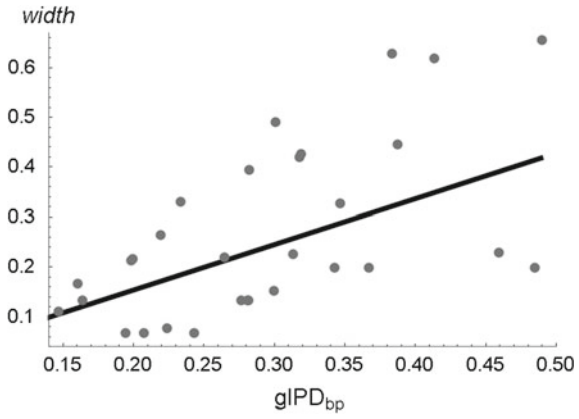


Fig. 24 Source width plotted over $g\mathrm{IPD}_{\mathrm{bp}}$ (*gray*) and the linear regression function (*black*)

As expected, the pair f_1 and I has a highly significant negative correlation. Six of the nine binaural quantities correlate significantly with width. The scatter and the linear regression function of $g\mathrm{IPD}_{\mathrm{bp}}$ are plotted in Fig. 24. 12 of the 15 binaural pairs also correlate significantly with each other, 8 of them on a $p < 0.01$ level. Most important for the multiple regression is the lower left region in the table. A pair of one monaural and one binaural sound field quantity is supposed to explain the source width. 3 monaural and 6 binaural quantities yield 18 potential pairs. However, 6 of them are ineligible, since they exhibit a significant correlation. Thus, they cannot be considered as orthogonal, which is a requirement for a valid multiple linear regression.

Results of multiple regressions with all pairs are summarized in Table 3. All 18 multiple regressions are significant ($p < 0.05$), 14 of them even highly significant ($p < 0.01$). Ineligible pairs that exhibit a correlation with each other are crossed out. Six of the combinations explain over 50 % of the variance, 5 of them are valid pairs. They are highlighted in gray. The linear combination of I and $g\mathrm{IPD}_{\mathrm{bp}}$

Table 3 Explained variance (R^2, top) and significance level (p-value) of multiple regressions between a pair of sound field quantities and source width

	ILD	gILD	gIPDbp	1—ICC	BPC	BEC
f_1	0.277 0.013	0.293 0.009	0.514 0.000058	~~0.55~~ ~~0.000021~~	0.484 0.000131	0.452 0.000295
I	~~0.331~~ ~~0.004~~	0.346 0.003	**0.615** **0.000002**	0.450 0.000315	~~0.497~~ ~~0.000093~~	~~0.471~~ ~~0.000187~~
B_{Rlog}	0.350 0.006	~~0.244~~ ~~0.035~~	0.512 0.000184	~~0.292~~ ~~0.016~~	0.588 0.000024	0.560 0.000052

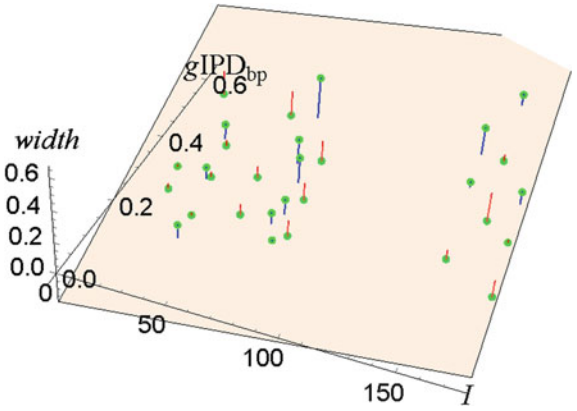


Fig. 25 Source width (*green*) plotted over I and $g\text{IPD}_{bp}$. The actual source width is connected to the predicted width which is based on multiple linear regression (transparent plane)

explains R^2 = 61.5 % ($p = 0.000002$) of the variance of source width. At an earlier state of research, R^2, the coefficient of determination, was 56 % ($p = 0.001601$) when considering only 8 instruments (Ziemer [50]). With a larger sample, including one inharmonic instrument, the results of the multiple linear regression improved. The result is illustrated in Fig. 25. Over-estimated widths are connected to the prediction plane with red lines, under-estimated widths with blue lines. It can be seen that the multiple linear regression yields a fair prediction of source width. This is even true for the extremes. No drastic outliers can be observed.

Some nonlinear combinations of I and $g\text{IPD}_{bp}$ yield slight improvements of the regression. Using the logarithm of the two, R^2 = 63.1 % of the variance is predictable, using their square root, R^2 becomes 63.2 %. A more effective nonlinear combination is similar to Eq. 7 as proposed by Ando [2], like

$$\text{ASW}_{pre} = aI^{1/3} + bg\text{IPD}_{bp}^{2/3} + c \tag{32}$$

which explained R^2 = 63.4 % of the variance.

5 Discussion

In this investigation, the radiation characteristics of 10 musical instruments has been measured. The radiated sound field is either directly measured at or propagated to 384 listening positions. Here, quantities from the field of psychoacoustics and subjective room acoustics have been calculated. Based on a pair of one monaural and one binaural parameter, the actual source width could be predicted with a fair precision. The best monaural predictor was the plain number of partials I in the considered frequency range. It is an even better predictor than the fundamental frequency or several measures of bass energy. Although the binaural pressure and energy components BPC and BEC exhibited a higher correlation with source extent, and even with a lower p-value, the weighted interaural phase difference below 1.2 kHz $g\mathrm{IPD}_{\mathrm{bp}}$ turned out to be the best predictor of source width, in combination with I.

This means that the number of partials might play a role in width perception. On the one hand, I is related to bass strength. The lower the fundamental frequency of musical instruments, the more partials in the spectrum tend to have an audible amplitude. From the literature, bass strength is already known to be related to the perception of source width. On the other hand, I is also closely related to spectral density. Spectral density might also be related to source extent and affect the perception of width.

Both versions of ILD significantly correlated with source width. This is in good agreement with the results derived from Potard and Burnett [39], that ILD are important for the recognition of shapes. It also seems to confirm the finding by Schroeder [43] that ILD are an important factor for a spatial sound impression. But $g\mathrm{IPD}_{\mathrm{bp}}$ gave the better prediction of width. This might imply that phase difference is an even more important parameter than level difference. This might be true in both a technical and a perceptual sense. It is interesting to see that a psychoacoustically motivated modification distinctly improved the results. A significant relationship could neither been found for IPD and width ($p = 0.289090$) nor between gIPD and width ($p = 0.114490$). But when considering only phase differences below the threshold of IPD perception, a high significance level is reached. This could mean that lower frequencies give more reliable cues for width perception. Of course, there are additional physical aspects: Considering a musical instrument as complex point source is a drastic simplification which is meaningful for low frequencies but it does not reflect the actual radiation characteristics of high frequencies well. Furthermore, due to the large wavelengths of low frequencies, slight misplacements of microphones hardly affect their measured phase. But for high frequencies, small misplacements can result in larger phase errors. As most of the considered partials lie above 1.2 kHz, the filtering eliminates these phase errors.

On the one hand, explaining 61.5 % of the variance is not very much. On the other hand, the number of considered instruments and listening distances is rather low. A higher R^2 is expected for a larger data set. This has proven to be true already: In an earlier state of this investigation, when only 8 instruments had been

measured, R^2 was 56 %. As even subjective judgments about perceived width provide a high variance, R^2 = 61.5 % might be sufficient for many applications. Considering and controlling the interaural phase differences of loud frequencies as well as the number of partials might be the right way to analyze and manipulate perceived source width. Of course, ICLDs and ICPDs in a stereo or surround setup do not create the same ILDs and IPDs. Zotter and Frank [54] have demonstrated that ICCC and IACC are proportional within a certain range. Naturally, ILD and IPD are lower than ICLD and ICPD. However, for a sweet spot, a simplified HRTF as proposed in Kling and Riggs [29] (p. 351) or a publicly available HRTF as published e.g. in Blauert et al. [7] and Algazi et al. [1] can be used to translate inter-channel differences to inter aural differences. In ambisonics and wave field synthesis systems where several listeners can move through an extended listening area, another method is necessary. One solution is to sample the listening area into a finite number of potential listening positions and create the desired $g\mathrm{IPD}_{\mathrm{bp}}$ here. This could be achieved by means of a high-order point multipole source as implemented in Corteel [12]. Alternatively, a rather coherent localization signal at each note onset is followed by the desired $g\mathrm{IPD}_{\mathrm{bp}}$ similar to the approach of Ziemer and Bader [53]. Likewise, DirAC encoding follows the idea to give one localization cue and one width cue. Such a coding could be used to give source position and $g\mathrm{IPD}_{\mathrm{bp}}$ as metadata.

6 Prospects

A reliable knowledge about the auditory perception of source width and the sound field at the listeners' ears is a powerful foundation for many applications. It could act as the basis of audio monitoring tools in recording studios to display perceived source width instead of plain channel correlations. This helps music producers to achieve the desired spatial impression. For channel-based audio systems, control over interaural cues is possible for a sweet spot if the loudspeaker positions are fixed and a HRTF is implemented. When using object-based audio coding, the desired interaural sound field quantities can be stored as metadata. This way, the approach can be adopted for a flexible use with arbitrary loudspeaker constellations. Instrument builders could focus on manipulating $g\mathrm{IPD}_{\mathrm{bp}}$ in a preferred listening region to achieve the desired perceived source extent. For example, the right radiation pattern could make a source sound narrow at one angle and more broad at another angle. Musical instruments for practicing could be designed to create a wider sound impression for the instrumentalist for a greater sound enjoyment. Then, instruments for performance create this sound impression for the audience. Simple measurement tools or advanced physical modeling software could support the work of instrument builders. Room auralization software can sound more realistic if it focuses on calculating the relevant parameters with high precision. Implementing radiation patterns of extended sources on sound field synthesis technologies, like

higher order ambisonics and wave front synthesis, can make the sound broader and more realistic. When concentrating on $g\mathrm{IPD}_{\mathrm{bp}}$ of partials as perceptually relevant parameters, computation time can be saved by synthesizing these cues instead of the whole radiation characteristics or other irrelevant parameters. This is again interesting for advancements in electric and electronic instruments. Electric pianos could sound more realistic, if the right auditory cues are recreated which make an actual grand piano sound this broad. Electric guitars could be widened and narrowed by turning one knob on the guitar amps which creates the desired monaural and interaural cues for a sweet spot or a limited listening region.

Until now, the presented approach lacks psychoacoustic proof. Listening tests under controlled conditions can bring reliable results concerning the relationship between sound radiation characteristics and perceived source extent. A prediction of source width may be more precise and especially more close to human perception when auditory processing is considered. Implementing binaural loudness and masking algorithms or even higher states of auditory processing is very promising to explain perceived source width in more detail.

References

1. Algazi, V.R., Duda, R.O., Thompson, D.M., Avendano, C.: The CIPIC HRTF database. In: IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New York, NY, pp. 99–102 (2001)
2. Ando, Y.: Auditory and Visual Sensation. Springer, New York (2010)
3. Baalman, M.: On Wave Field Synthesis and Electro-acoustic Music, with a Particular Focus on the Reproduction of Arbitrarily Shaped Sound Sources. VDM, Saarbrücken (2008)
4. Beranek, L.L.: Concert Halls and Opera Houses: Music, Acoustics, and Architecture, 2nd edn. Springer, New York (2004)
5. Blau, M.: Correlation of apparent source width with objective measures in synthetic sound fields. Acta Acust. United Acust **90**(4), 720–730 (2004)
6. Blauert, J.: Spatial Hearing. The Psychophysics of Human Sound Source Localization (Revised edn.). MIT Press, Cambridge (1997)
7. Blauert, J., Brüggen, M., Hartung, K., Bronkhorst, A.W., Drullmann, R., Reynaud, G., Pellieux, L., Krebber, W., Sottek, R.: The AUDIS catalog of human HRTFs. In: Proceedings of the 16th International Congress on Acoustics, vol. 4, pp. 2901–2902, Seattle (1998)
8. Blauert, J., Cobben, W.: Some consideration of binaural cross correlation analysis. Acta Acust. United Acust **39**(2), 96–104 (1978)
9. Blauert, J., Lindemann, W.: Auditory spaciousness: some further psychoacoustic analyses. J. Acoust. Soc. Am. **80**(2), 533–542 (1986)
10. Bradley, J.S., Reich, R.D., Norcross, S.G.: On the combined effects of early- and late-arriving sound on spatial impression in concert halls. J. Acoust. Soc. Am. **108**(2), 651–661 (2000)
11. Cabrera, A.: Pseudo-stereo techniques. CSound implementations. Csound J. **14** (Article number 3) (2011)
12. Corteel, E.: Synthesis of directional sources using wave field synthesis, possibilities, and limitations. EURASIP J. Adv. Sign. Process. Article ID 90509 (2007)
13. Damaske, P., Ando, Y.: Interaural crosscorrelation for multichannel loudspeaker reproduction. Acta Acust. United Acust **27**(4), 232–238 (1972)

14. de Vries, D., Hulsebos, E.M., Baan, J.: Spatial fluctuations in measures for spaciousness. J. Acoust. Soc. Am. **110**(2), 947–954 (2001)
15. Deutsches Institut für Normung.: Akustik — Messung von Parametern der Raumakustik — Teil 1. Aufführungsräume (ISO 3382-1:2009); Deutsche Fassung EN ISO 3382-1:2009 (2009)
16. Faller, C.: Pseudostereophony revisited. In: 118th Audio Engineering Society Convention, Barcelona (2005)
17. Friedrich, H.J.: Tontechnik für Mediengestalter. Töne hören — Technik verstehen — Medien gestalten. Springer, Berlin (2008)
18. Gade, A.C.: Acoustics in halls for speech and music. In: Rossing, T.D. (ed.) Handbook of Acoustics, Chapter 9, pp. 301–350. Springer, Berlin (2007)
19. Gerzon, M.A.: The design of precisely coincident microphone arrays for stereo and surround sound. In: 50th Audio Engineering Society Convention, London (1975)
20. Griesinger, D.: Objective measures of spaciousness and envelopment. In: AES 16th International Conference: Spatial Sound Reproduction, Rovaniemi (1999)
21. Haas, H.: Einfluss eines Einfachechos auf die Hörsamkeit von Sprache. Acustica **1**, 49–58 (1951)
22. Hamidovic, E.: The Systematic Mixing Guide. Systematic Productions, Melbourne (2012)
23. Heller, A.J.: Is my decoder ambisonic? In: 125th Audio Engineering Society Convention, San Francisco, CA (2008)
24. Hirvonen, T., Pulkki, V.: Center and spatial extent of auditory events as caused by multiple sound sources in frequency-dependent directions. Acta Acust. United Acust. **92**(2), 320–330 (2006)
25. Jacques, R., Albrecht, B., Melchior, F., de Vries, D.: An approach for multichannel recording and reproduction of a sound source directivity. In: 119th Audio Engineering Society Convention, New York (2005)
26. Kaiser, C.: 1001 Mixing Tipps. mitp, Heidelberg (2012a)
27. Kaiser, C.: 1001 Recording Tipps. mitp, Heidelberg (2012b)
28. Kaiser, C.: 1001 Mastering Tipps. mitp, Heidelberg (2013)
29. Kling, J.W., Riggs, L.A. (eds.): Woodworth & Schlossberg's Experimental Psychology, 3rd edn. Holt, Rinehart and Winston, New York (1971)
30. Laitinen, M.-V., Philajamäki, T., Erkut, C., Pulkki, V.: Parametric time-frequency representation of spatial sound in virtual worlds. ACM Trans. Appl. Percept. **9**(2) (2012)
31. Laitinen, M.-V., Walther, A., Plogsties, J., Pulkki, V.: Auditory distance rendering using a standard 5.1 loudspeaker layout. In: 139th Audio Engineering Society Convention, New York, NY (2015)
32. Levinit, D.J.: Instrument (and vocal) recording tips and tricks. In: Greenbaum, K., Barzel, R. (eds.) Audio Anecdotes, vol. I, pp. 147–158. A K Peters, Natick (2004)
33. Lindemann, W.: Extension of a binaural cross-correlation model by contralateral inhibition. ii. the law of the first wave front. J. Acoust. Soc. Am. **80**(6), 1623–1630 (1986)
34. Maempel, H.-J. (2008). Medien und Klangästhetik. In: Bruhn, H., Kopiez, R., Lehmann, A.C. (eds.) Musikpsychologie. Das neue Handbuch, pp. 231–252. Rowohlt, Reinbek bei Hamburg (2008)
35. Martín, R.S., Witew, I.B., Arana, M., Vorländer, M.: Influence of the source orientation on the measurement of acoustic parameters. Acta Acust. United Acust. **93**(3), 387–397 (2007)
36. Mason, R., Brookes, T., Rumsey, F.: The effect of various source signal properties on measurements of the interaural crosscorrelation coefficient. Acoust. Sci. Technol. **26**(2), 102–113 (2005)
37. Okano, T., Beranek, L.L., Hidaka, T.: Relations among interaural cross-correlation coefficient ($IACC_E$), lateral fraction (LF_E), and apparent source width (ASW) in concert halls. J. Acoust. Soc. Am. **104**(1), 255–265 (1998)
38. Otondo, F., Rindel, J.H.: The influence of the directivity of musical instrument in a room. Acta Acust. United Acust **90**, 1178–1184 (2004)
39. Potard, G., Burnett, I.: A study on sound source apparent source shape and wideness. In: Proceedings of the 2003 International Conference on Auditory Display, Boston, MA (2003)

40. Potard, G., Burnett, I.: Decorrelation techniques for the rendering of apparent sound source width in 3d audio displays. In: Proceedings of the 7th International Conference of Digital Audio Effects, Naples (2004)
41. Rogers, S.E.: The art and craft of song mixing. In: Greenbaum, K., Barzel, R. (eds.) Audio Anecdotes, vol. II, pp. 29–38. A K Peters, Natick (2004)
42. Ross, B., Tremblay, K.L., Picton, T.W.: Physiological detection of interaural phase differences. J. Acoust. Soc. Am. **121**(2), 1017–1027 (2007)
43. Schroeder, M.R.: An artificial stereophonic effect obtained from using a single signal. In: 9th Audio Engineering Society Convention, New York, NY (1957)
44. Shimokura, R., Tronchin, L., Cocchi, A., Soeta, Y.: Subjective diffuseness of music signals convolved with binaural impulse responses. J. Sound Vibr. **330**, 3526–3537 (2011)
45. Slavik, K.M., Weinzierl, S.: Wiedergabeverfahren. In: Weinzierl, S. (ed.) Handbuch der Audiotechnik, Chapter 11, pp. 609–686. Springer, Berlin (2008)
46. Yanagawa, H., Anazawa, T., Itow, T.: Interaural correlation coefficients and their relation to the perception of subjective diffuseness. Acta Acust. United Acust. **71**(3), 230–232 (1990)
47. Yanagawa, H., Tohyama, M.: Sound image broadening by a single reflection considering temporal change of interaural cross-correlation. Acta Acust. United Acust. **87**(2), 247–252 (2001)
48. Yanagawa, H., Yamasaki, Y., Itow, T.: Effect of transient signal length on cross-correlation functions in a room. J. Acoust. Soc. Am. **84**(5), 1728–1733 (1988)
49. Ziemer, T.: Sound radiation characteristics of a shakuhachi with different playing techniques. In: Proceedings of the International Symposium on Musical Acoustics, Le Mans, pp. 549–555 (2014)
50. Ziemer, T.: Adapting room acoustic parameters to explain apparent source width of direct sound. In: 'Musik und Wohlbefinden'. 31. Jahrestagung der DGM, Oldenburg, pp. 40–41 (2015)
51. Ziemer, T.: Wave field synthesis. In: Handbook of Systematic Musicology. Springer, Berlin (in Print) (2016)
52. Ziemer, T., Bader, R.: Complex point source model to calculate the sound field radiated from musical instruments. In: Proceedings of Meetings on Acoustics, vol. 25 (2015a)
53. Ziemer, T., Bader, R.: Implementing the radiation characteristics of musical instruments in a psychoacoustic sound field synthesis system. In: 139th Audio Engineering Society Convention, New York, NY (2015b)
54. Zotter, F., Frank, M.: Efficient phantom source widening. Arch. Acoust. **38**(1), 27–37 (2013)
55. Zotter, F., Frank, M., Kronlachner, M., Choi, J.-W.: Efficient phantom source widening and diffuseness in ambisonics. In: Proceedings of the EAA Joint Symposium on Auralization and Ambisonics, Berlin (2014)
56. Zwicker, E., Fastl, H.: Psychoacoustics. Facts and Models (Second updated edn.). Springer, Berlin (1999)

Author Biography

Dr. Tim Ziemer is a musicologist, mainly working in the field of applied psychoacoustics from wave field synthesis over music information retrieval to instrument acoustics. His research interests include the perception of spaciousness and its relation to sound radiation characteristics of musical instruments. Tim Ziemer has worked with research teams at the University of Hamburg and the National Institute of Informatics Tokyo. He is a freelance author for a renowned computer magazine and has a professional background in teaching, music production, room and building acoustics, concert logistics and cultural administration.

Methods in Neuromusicology: Principles, Trends, Examples and the Pros and Cons

Christiane Neuhaus

Abstract Neuromusicology, also known as the Cognitive Neuroscience of Music, is a modern discipline devoted to the measurement of real-time processes in the human brain while perceiving and producing sound. Research topics range from acoustic feature processing and listening to melodies to composition and music performance. Before designing an experiment, researchers might find it helpful to be informed about the efficiency of methods and their pros and cons. The chapter at hand gives an overview of several methods used in the neurosciences with a special emphasis on their principles, constraints and fields of application. The focus is on transcranial magnetic stimulation (TMS), functional magnetic resonance imaging (fMRI), positron emission tomography (PET), electroencephalography (EEG) and on event-related potentials (ERP). The reader will also become acquainted with trends and recent developments towards whole-brain analyses and real life studies based on the idea to improve ecological validity.

1 Introduction

Neuromusicology, also termed the 'Cognitive Neuroscience of Music', is a modern discipline that came into existence through its methods. It is still not clear whether it belongs to cognitive neuroscience as a sort of 'parent discipline', to empirical musicology or whether it has a status of its own. Any neuroscience method enables researchers to measure the brain's physiological processes in real-time, thus giving insight into the task-related or spontaneous functionings of the human brain without requiring any verbal or behavioral type of response.

The chapter gives an overview about the most frequent methods used in neuromusicology. I will particularly focus on transcranial magnetic stimulation (TMS), functional magnetic resonance imaging (fMRI), positron emission tomography (PET), electroencephalography (EEG) and on event-related potentials (ERP). These

C. Neuhaus (✉)
Institute of Systematic Musicology, University of Hamburg, Hamburg, Germany
e-mail: christiane.neuhaus@uni-hamburg.de

A. Schneider (ed.), *Studies in Musical Acoustics and Psychoacoustics*,
Current Research in Systematic Musicology 4,
DOI 10.1007/978-3-319-47292-8_11

types of research methods are all non-invasive, i.e. neurosurgical interventions are circumvented. For each method I will weigh up the pros and cons and give examples of application.

Since results obtained with these techniques essentially differ in nature, it is necessary to first classify methods according to type: Data achieved with EEG and ERP belong to the class of 'bioelectric potentials'. Those achieved with fMRI and PET belong to the group of 'neuroimaging data'. Results obtained with EEG and ERP have shifts of intracellular currents as their common source (and starting point), and bioelectric potentials on the head's surface are the final output (for details regarding electrogenesis, see Sect. 3). Thus, by using EEG and ERP, neural activity can *directly* be measured. Scans obtained with fMRI and PET, by contrast, reveal changes in energy consumption (mostly oxygen) in well-defined regions of the brain. These types of neuroimaging techniques therefore point to neural activity in a merely *indirect* way.

Jäncke [13, 14] addressed this issue of method classification in a slightly different manner by using the exact reference to anatomical structures as a specifying criterion. According to that, neuroimaging methods like fMRI and PET allow a precise assignment of physiological processes to neuroanatomical structures, whereas bioelectrical methods like EEG or ERP do not.

1.1 Transcranial Magnetic Stimulation: How Does It Work?

Let me start with some essential remarks on transcranial magnetic stimulation (TMS). In comparison to the methods mentioned above TMS is a sort of exception: It enables researchers to draw certain *causality-based* conclusions, i.e. to precisely relate 'cause' and 'effect'. In contrast to that, fMRI, ERP and other conventional neuroscience methods, although quite popular, allow only mutual relationships, or *correlative conjunctions,* which means that coincidences cannot be ruled out.

How does TMS work? Transcranial magnetic stimulation modifies the excitability of nerve tissue so that cortical processes may either be accelerated or severely be inhibited up to virtual lesions between 10 and 30 min length. The underlying principle is electromagnetic induction: A magnetic field is temporarily built up in orthogonal direction to the plane of a stimulation coil that is placed 2 cm above the head. This way, an electric current is induced in the small cortical regions underneath while tissue resistances of skin, skull and dura can be disregarded [13]. It depends on the coil's stimulation frequency whether processes speed up or slow down: A pulse series with repetition rates of 5 Hz or higher (repetitive TMS; rTMS) may cause facilitation through lowering the excitation threshold, whereas a pulse series with repetition rates of 1 Hz maximum (same for 10-ms intense single shots) may have the opposite effect and provoke inhibition by suppressing the intracellular current flows [13].

Note that, in principle, TMS does not belong to the class of 'neuroimaging or visualization methods'. Instead, TMS modulates pure neurophysiological (excitatory and inhibitory) activity, observable either through a decelerated or an accelerated overt behavior [13]. However, to complete the data, researchers often decide for a combined approach using TMS as well as fMRI (or ERP) for measuring purposes (e.g. [2]).

Things are more complicated regarding the auditory cortex: During stimulation, abrupt electromagnetic forces make the TMS coil produce noise, causing severe, disruptive effects, in particular when early auditory processes are investigated. Jäncke [13] reported that old apparatus produce sharp coil noise between 120 and 130 dB intensity (which is near the auditory pain threshold), making the use of earplugs necessary. In a combined high-resolution ERP-TMS study Tiitinen et al. [57] tested TMS noise of three volume intensities (80, 90 and 100 dB SPL) for its distorting effects. The study shows that coil clicks alone evoked auditory brain responses with a maximum amplitude at 170 ms post stimulus-onset. Furthermore, clicks seem to interact with a series of simultaneously presented pure tones serving as target stimuli (1000 Hz, 85 dB SPL, 50 ms duration, inattentively processed while reading a book), leading to attenuated brain responses in terms of the target tones. Due to these possible effects of contamination, TMS is less often considered as the method of choice whenever pure listening has to be investigated. However, by using elegant paradigms some difficulties in auditory rTMS research can be avoided: Andoh and Zatorre [2], for example, approached this issue by disentangling the time span between coil stimulation and subsequent testing: Two types of rTMS sequences were initially presented off-line in that the auditory cortex was first stimulated with trains of one pulse per second during the first 10 min, and then with a volley of ten pulses per second (followed by an ITI of 10 s) over the next 26 min, totaling up to 1200 (2×600) pulses (Fig. 1). Immediately after stimulation, a melody discrimination task had to be solved, showing interesting sex- and time-related results depending on the type of stimulation: Female participants (8 males, 8 females) significantly accelerated their recognition performance after stimulation with 10 Hz-rTMS sequences (RT became shorter), whereas male participants showed the opposite (RT became longer). In the second half of testing, the situation went into reverse, and for the female group, processing slowed down again.

2 Functional Magnetic Resonance Imaging: Basic Principles and Image Acquisition

Let me move on to the conventional types of neuroscience methods. I will first take a closer look at the principles of functional magnetic resonance imaging (fMRI). Since the underlying mechanisms are quite complex, I will omit the main parts of MR physics here and restrict myself to those aspects I consider relevant for

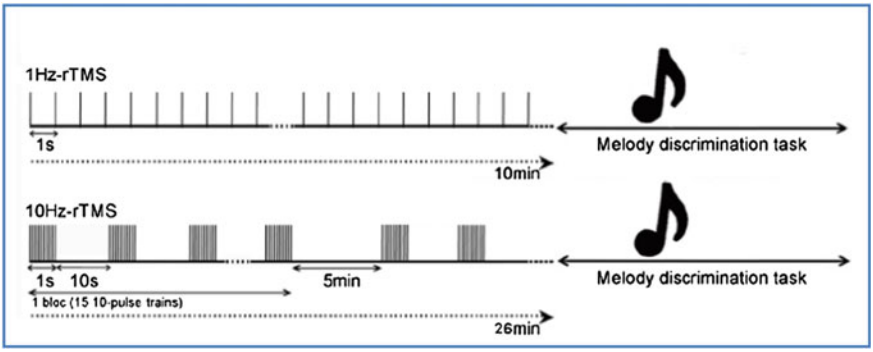


Fig. 1 Repetitive TMS (rTMS): for stimulating brain tissue different types of pulse sequences are in use [2]

discussing the methods' advantages and disadvantages. (For further information on MR physics please refer to the detailed, excellent explanations in [13, 19].

To investigate brain activity with fMRI, measurement has to be performed in two steps: In the initial phase, called Magnetic Resonance Imaging (MRI), pure neuroanatomical data are recorded to precisely reconstruct size, shape and structure of the individual brain. After that, changes of regional blood flow are registered to determine the amount of oxygen consumption. (This is the method's functional component, the 'f' in **f**MRI). Regarding the equipment, an expensive magnetic resonance scanner (tomograph) is the indispensable part; it should be able to produce a high-intense, static magnetic field B_0 with a field strength of 1 up to 7 Tesla (T).

MR physics uses the spin, i.e. the self-rotating property of hydrogen atoms (H). Inside the scanner, the hydrogen atoms of the human body (4.7×10^{27} in number) react like tiny compass needles. They orient along the external magnetic field B_0 and rotate with a special frequency termed 'Larmor precession frequency'. Whenever B_0 is modulated by introducing a sharp HF-impulse hydrogen atoms fold down from the Z- into the XY-plane, and magnetization changes from longitudinal to diagonal. To re-reach the starting position two types of 'backward forces', called 'spin-grid interaction' and 'spin-spin interaction' are effective, widely known as T1- and T2-relaxation [13].

In physical terms, T1-relaxation is defined as the point in time when longitudinal magnetization has regained 63 % of its original strength, whereas T2-relaxation is defined as the point in time when diagonal magnetization has decreased to 37 % of its initial value [13]. Gray and white matter, fat and cerebrospinal liquid differ in their relaxation times (T1 and T2), which enables researchers to use these parameters for adjustment, in particular for regulating brightness and image contrast of the MR scans. Neuromusicology uses almost exclusively MR scans of the T1-weighed type. This type of brain image makes gray and white matter (as well as other forms

of tissue with short T1-properties) look bright, whereas spaces filled with cerebrospinal liquid look dark, producing almost no signal.

In addition, several sub-processes, known as space-coding, are necessary to obtain precise 2- and 3-dimensional *spatial information* from the spin signal. Procedures for space-coding require a stimulation of selective layers as well as a modulation of the magnetic field and for this, paired gradient coils of X-, Y-, and Z-orientation are used [13]. These coils produce gradient fields superimposed on B_0, resulting in different strengths of the magnetic field at each point of the measuring volume. Regarding spatial information the MR signal can now be analyzed voxel by voxel [49].

Functional MR scans are the product of a further step of image acquisition, and for this aim, the so-called BOLD response is a necessary precondition (see next chapter).

2.1 BOLD Response and Its Underlying Principle

Let me describe the *physiological mechanism* responsible for the '**f**' in **f**MRI: Blood itself, more precisely, the oxygen content of hemoglobin serves as a body-own (intravascular) indicator: Whenever a task has to be performed, either motor or cognitive, energy demand is high, and the regional cerebral blood flow (rCBF) increases. In the venous parts of the fine capillaries next to activated neural populations [49] an exchange of oxygen-deficient for oxygen-rich blood takes place, and this is the principle fMRI is based on. Note that the oxygenated type of hemoglobin differs from the de(s)oxygenated one in its magnetic susceptibility, the latter being slightly more susceptible than the former (also known as the para- vs. diamagnetic properties of blood; [13]). The principle of magnetic susceptibility was originally discovered by the American chemist Linus Pauling in the 1920s and was transferred to fMRI research in 1990, since then known as BOLD response (**B**lood **O**xygen **L**evel **D**ependent response, [36]). Jäncke [13] explains that deoxygenated (paramagnetic) hemoglobin has an inhibiting effect on the BOLD signal due to increasing magnetic field inhomogeneities yielding artifacts, but he concedes that the underlying mechanisms of signal production are still more complex.

The BOLD signal usually reaches its peak plateau between 4 and 8 s after task-onset. Thus, in comparison to bioelectrical methods, time resolution is poor. However, the main advantage of fMRI lies in its excellent spatial resolution with values between 1 and 4 mm^3 (or even below) which is a necessary precondition to precisely localize functionally activated areas.

Note that Talairach and Tournoux [53], two neurosurgeons of Swiss origin, developed a stereotaxic atlas, i.e. a sort of spatial coordinate system supporting researchers in their effort to localize specific brain areas. By using a color range from red to blue to indicate either activation or deactivation, clusters of functional activity can be marked voxel- or pixelwise on these brain scans. This stereotaxic atlas also provides the possibility to adjust the morphological structure of individual

brains or to transform data onto a template, the standard brain, used for data transfer between laboratories and for the comparability of results.

2.2 Techniques of Image Acquisition

Constructing an appropriate paradigm in fMRI research is no easy matter:

First, several options of pulse sequences must prove their suitability to stimulate the brain tissue in an adequate manner. Three of them, called *spin-echo-sequence*, *gradient-echo-sequence* and *echoplanar imaging (EPI)*, will be briefly introduced here (Fig. 2 shows similar considerations regarding TMS). In principle, stimulation starts with a HF-pulse, then a special combination of gradient coils is applied [13]. *Spin-echo-sequences* simply work with two (or more) HF-pulses. The initial 90° HF-pulse makes hydrogen atoms fold down into the XY-plane, whereas a second 180° HF-pulse forces the atoms to reverse their direction of precession; resulting in maximal strength of the to-be-analyzed signal. *Gradient-echo-sequences* and spin-echo-sequences differ from each other in a crucial point: Gradient coils change polarization on their own, which makes the process of signal generation less time-consuming while image quality of T1-weighed scans remains excellent.

Fig. 2 Example of a T1-weighed scan (Universidad Autónoma de Zacatecas, Mexico)

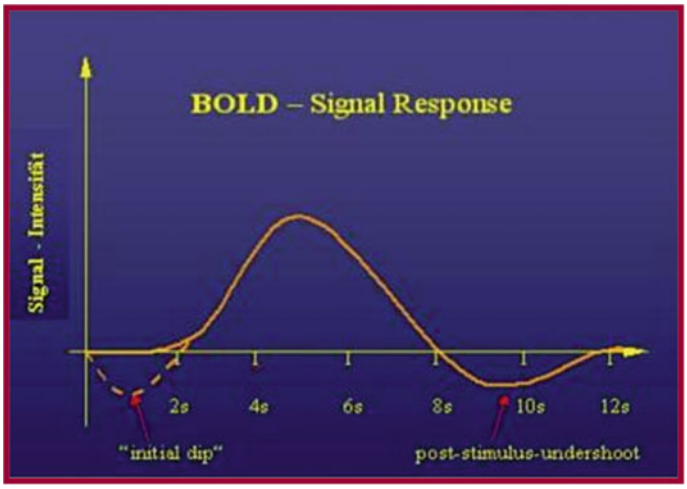


Fig. 3 Time course of a BOLD response [49]

Echoplanar imaging (EPI) is the fast, high-resolution version of the gradient-echo technique. EPI enables researchers to record the whole brain in less than 2 s. A disadvantage of this stimulation type is, again, coil noise: The fast and permanent switching of gradient coils transmits small amounts of vibration onto the cylindrical scanner tube which automatically starts to resonate by producing volume intensities between 60 and 100 dB. Thus, it seems advisable to wear protective headphones or earplugs.

A second constraint is related to the BOLD signal itself: The fact that it needs at least 4 s to reach its plateau and, after decay, needs 12 to 15 s to recreate, imposes severe restrictions on the experimental design, meaning that signal presentation and inter-trial intervals (ITIs) have to be adapted to these limitations (Fig. 3).

2.3 The Auditory Cortex—A Challenge to fMRI Research

About 87 % of all fMRI studies use echoplanar imaging (EPI) as the method of choice, valued for rapid data collection [29]. Investigating the auditory cortex with fMRI, however, is a special problem. As already mentioned above, a disadvantage is that noise is produced by a fast switching of gradient coils during space-coding, ranging from 60 to 100 dB intensity. This side effect, similar to that observed for transcranial magnetic stimulation (TMS), makes the study of auditory processing complicated for, at least, three reasons: First, target sounds could be masked by the scanner noise and may hardly be detected. Second, the focus of attention may shift towards the noise and will eventually impair task performance. Third, emotional reactions to melodies as well as their aesthetic evaluation might extremely be

hindered, resulting in reduced brain activation of the limbic system (amygdala and hippocampus).

How can researchers deal with these side effects caused by scanner disturbance? Several effective solutions are suggested by Jäncke [13]: The first compensating technique he introduces is known as *sparse temporal sampling (STS).* It is a variation of the 'echo-planar imaging' technique. STS is characterized by inserting pauses of up to 10 s length into periods of continuous scanning. These inter-stimulus intervals can make the coil noise fade out while at the same time, i.e. preceding the measurement, target melodies may fade in. Another possibility is termed *clustered acquisition*. This time, the entire set of target stimuli will be presented immediately before scanning, providing a time-frame between 4 and 6 s for image acquisition (data recording). Note that whenever *clustered acquisition* is used as the method of choice, the first two (of let's say, ten) fMRI scans have to be excluded from analysis: Owing to the fact that longitudinal magnetization has not been completely built up, signal strength is less than in the remaining scans [13]. Several other simple measures may also effectively reduce scanner noise. First, scanner-compatible headphones may attenuate the coil noise while at the same time the sound quality of target melodies can be enhanced. Another possibility is to line the inside of the scanner tube with sound-absorbing material (e.g. insulating mats) with attenuation effects up to 30 dB.

Faced with the challenge to increase the effectiveness of auditory fMRI designs Mueller et al. [31] tested the effects of three types of EPI-sequences on image acquisition: *continuous scanning*, *sparse temporal sampling (STS)* as well as a method called *interleaved silent steady state (ISSS)* that differs from STS in terms of scanner behavior in the silent gaps between recordings.

Mueller et al. [31] tested all three types of EPI sequences (continuous scanning, STS and ISSS) for possible differences in brain activation caused by the measuring technique itself. In each session of 12.5 min length the same 10 s excerpts of classical dances and their electronically distorted counterparts were taken as example pieces to examine the excitability of brain tissue in 20 volunteers (7 females). The study obtained two results: First, activations in left and right auditory cortices were significantly stronger for the original than for the distorted dance pieces. More interesting, however, is the observation of additional activations in the limbic system (left and right hippocampal structures) that could be made visible with ISSS but not with sparse temporal sampling. So, unexpectedly, the interleaved silent steady state method emerged as the most sensitive acquisition technique; thus, it might be the method of choice whenever subtle activities in subcortical structures or in deeper layers of the cerebrum have to be detected.

2.4 Positron Emission Tomography: Some Notes on the Signal and on Image Acquisition

It is the right time to take a closer look at positron emission tomography (PET), the older type of neuroimaging methods used for visualizing brain processes. Once again, energy consumption (oxygen and glucose) serves as an indicator to precisely localize brain activity. Spatial resolution obtained with this method is between 2 and 8 mm^3 depending on the type of PET scanner, i.e. it is slightly less accurate than that obtained with fMRI. When thinking about auditory paradigms neuroscientists often choose PET instead of fMRI. The reason why PET is preferred lies in the avoidance of scanner noise due to a different technique of data recording. In other words, PET in contrast to fMRI does not produce disturbing scanner noise at all. Thus, while lying in the tube of a PET scanner, participants can easily focus on the target sound, react emotionally and may also appreciate the target's aesthetic value. However, the major disadvantage of this method is that a radioactive tracer substance, mostly 15-Oxygen (^{15}O) or 2-Fluoro-2-Deoxyglucose, has to be injected intravenously. Radioactive isotopes emit positrons (particles of positive electric charge) that interact (or collide) with electrons (particles of negative electric charge), resulting in the emission of photons which can be measured with an array of scintillation detectors, i.e. a circle of sensors placed around the head [13].

Note that the half-life of each radioisotope imposes restrictions on the experimental paradigm in that time length available for task-related measurement is strictly determined by the rate of decay. The half-life of 15-Oxygen, for instance, is about 2 min, placing severe restrictions on the choice of appropriate stimuli: Single sound events such as pure tones or intervals have been proven suitable, whereas harmonic sequences or melodies should not exceed a time length of, on average, 10 s.

Jäncke [13] points out that due to risks of health only 10 injections of radioisotopes per participant seem acceptable, resulting in 20 min (10 shots à 2 min each) recording time, also restricting the number of the to-be-tested conditions. In compensation, and also to increase the statistical power, group-wise averaging of PET scans is advisable, and for this, PET data first have to be transferred to the template, the standard brain. Thus, in terms of *individual* PET data it seems almost impossible to achieve convincing statistical results.

2.5 Research with FMRI and PET: Example Fields of Music-Related Application

Scans depict functional activity in specific brain regions, using energy consumption as an indicator. Most results refer to the cerebrum, in particular the cortex, the basal ganglia and the limbic system, but, increasingly, the cerebellum and parts of the brain stem down to the pons (in particular the cochlear nucleus and the superior

olive of the *auditory* brain stem) are also investigated using neuroimaging methods (e.g. [50]). To give you a glimpse of an idea about the range of results obtained with fMRI and PET I will pick some out, choosing 'auditory processing', 'music theory' and 'creativity' as example fields of application.

Note, however, that from a neurophilosophical perspective fMRI results do not allow to draw a conclusion about the functioning of the *mind* or the type of knowledge representation (be it analogous or propositional; see e.g. [41, 45] for a further discussion). In other words: neuroimaging methods still cannot be used to distinguish between mind and brain, the old, 'hard' philosophical problem. Even so, first attempts have been made to reconstruct the mental content belonging to different semantic categories from brain scans showing cortical activation and deactivation (see e.g. [12]).

2.5.1 Studying the Human Auditory Cortex with PET and FMRI

Neuroimaging methods enable researchers to examine the specific functioning of the auditory cortex in detail. This way, many fundamental insights that were previously found by introspection, now can be verified via brain scans which may help disciplines like *Psychophysics* and *Tonpsychologie* strengthen their impact.

One example in this respect is a study performed by Zatorre and Belin [60]. Using PET, they were able to identify two functionally different parts in the auditory cortex, a core and a belt region. The core region is specialized in processing temporal features as typical for speech, whereas the belt region is specialized in processing spectral features as typical for tonal patterns. Furthermore, they observed a certain asymmetric shift (or functional lateralization) in that speech-like signals caused stronger activation in the left compared to the right auditory cortex whereas for signals with rich spectral content as in music the reverse was true. Zatorre and Belin [60] suggest that neuroanatomical structures on the micro-level, in particular different types of fiber myelination and cortical column width, may be the reason why rapid changes of signals are processed in the left auditory cortex, whereas spectral richness is more accurately processed in the right counterpart. (Note: It is the idea of starting with a common signal which was split up in two directions that makes the study trustworthy: Two pure tones in octave distance served as the standard signal and were randomly presented in alternating order. To either synthesize speech signals or simulate music, they were then speeded up in three steps (first condition) and enriched with additional spectral components (second condition)).

In addition, Warren et al. [59] could show by using fMRI that two regions of the secondary auditory cortex, termed planum polare (PP) and planum temporale (PT), react independently to pitch chroma and pitch height, the cyclical and linear dimensions of pitch as postulated by Drobisch [6] and by Révész [42]; (see Fig. 4). These findings confirm that the so-called 'Two-component theory of pitch' (*Zweikomponententheorie der Tonhöhe*), which is originally based on subjective assessments, indeed has a neuroanatomical counterpart. (To obtain these results,

individual sounds were manipulated in pitch height by attenuating the amplitude of the odd harmonics while keeping pitch chroma constant).

2.5.2 Tonality-Sensitive Areas—An Approach with fMRI

Another example field of application is music theory, although publications about the general laws in music and their neural base are scarce. Yet, Janata et al. [16] discovered a sort of *tonality-sensitive center* located in the rostromedial part of the prefrontal cortex (rmPFC). In this fMRI study, a melody was systematically modulated through all 24 major and minor keys, and eight musically experienced participants listened attentively to each transposed version.

While comparing the congruency of cortical activation patterns between transpositions, Janata et al. observed a certain "dynamic topography" in rmPFC (p. 2167) in that some activated voxels specifically responded to a certain group of keys but not to another. However, repetitions of the scanning procedure revealed that inter- as well as intraindividual variances in terms of these *key-sensitive* voxel-based activations were high (Fig. 5). Note that this type of key-sensitive, dynamic reaction in rmPFC was independent of the given task, (consisting in detecting tones of different timbre as well as in detecting those ones being out-of-key).

2.5.3 Musical Improvisation—An Example of Whole-Brain Image Analysis

In recent years, a tendency towards naturalness and authenticity has been observed showing that 'high ecological validity' becomes a core criterion in cognitive

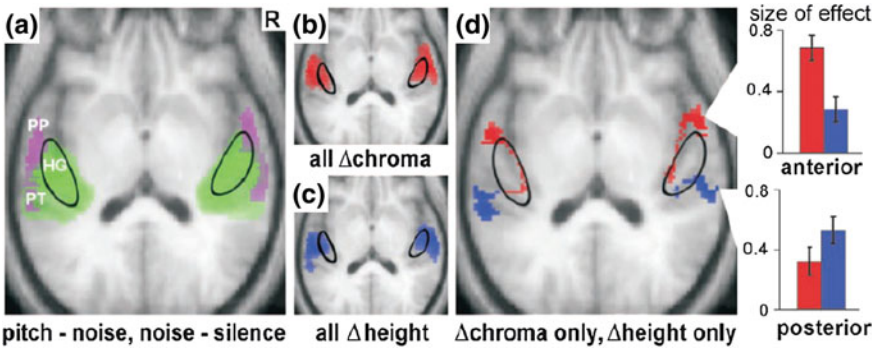


Fig. 4 Planum polare (PP) and planum temporale (PT) are differently activated when pitch height and pitch chroma are processed independently, confirming the 'two-component theory of pitch' [59]

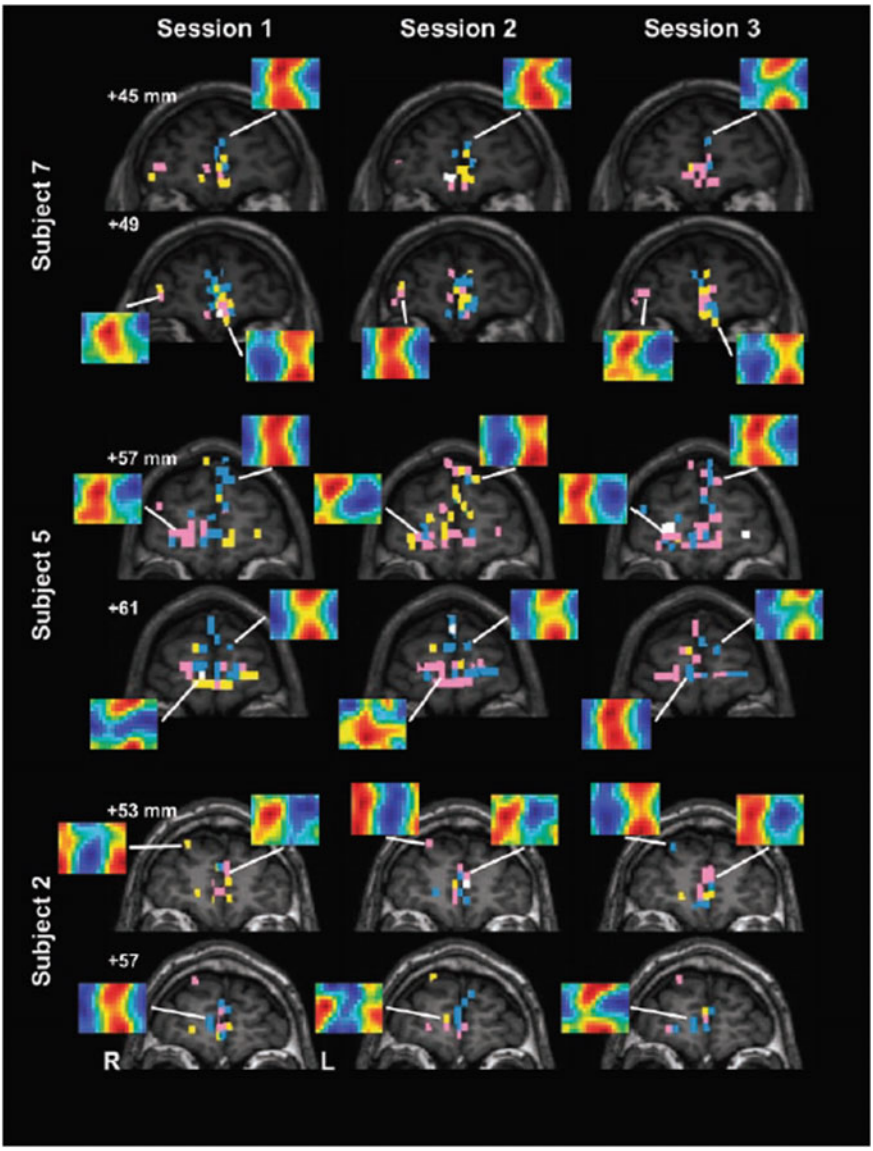


Fig. 5 Brain scans of subjects 2, 5 and 7 across three scanning sessions: the rostromedial PFC is sensitive to tonality. Within this area certain activated voxels reveal a key-specific behavior [16]

neuroscience. In this context some *neuromusicological field studies* (using EEG) occasionally have been put into practice: Fritz and colleagues, for instance, used a portable EEG equipment to record brain activity from native village inhabitants in Cameroon who listened to Western classical music for the first time (unpublished

work). A mobile EEG equipment has also been found useful to examine the effects of Cannabis on consuming rock music while sitting in the living room, smoking a couple of joints (see [9]).

Neuroimaging methods, by contrast, cannot fully meet the criterion of context-related authenticity as the scanning procedure should always be performed in a laboratory environment to obtain trustworthy results. Despite these constraints, a new tendency in fMRI research can be observed in that complex, natural musical pieces of several minutes length are used to investigate free natural listening and/or some types of spontaneous creativity.

In this context, Alluri et al. [1] demonstrated with fMRI that free listening to a modern tango of 8′32″ length activated cognitive, motor and emotion-related circuits on cortical and subcortical levels simultaneously while, at the same time, deactivations, mostly in pre- and postcentral cortical areas, were found (Fig. 6). In order to reliably relate the respective brain activity to the musical parameters of this tango, 21 music students judged 270 tango segments of 6 s length on a scale from 1 to 9 beforehand, according to the following parameters: 'brightness', 'fullness', 'activity' (i.e. 'change of timbre'), 'timbral complexity', 'rhythmic complexity', 'pulse clarity', 'key clarity', 'event synchronicity' and 'dissonance'. Interestingly, timbre features were not only processed in the auditory cortices, but also in some parts of the left and right cerebellum, while, at the same time, deactivations in the precentral and parietal regions could be found. 'Pulse clarity' was processed in the auditory cortices too, but showed deactivations in the insula and the limbic system (amygdala, hippocampus; see Fig. 6).

Complexity can even be increased when creative and performance aspects are added. One of the first neuroimaging studies addressing this issue was an fMRI experiment by Limb and Braun [28], testing the neural basis of improvisation. In regions of the prefrontal cortex, a dissociated pattern of activity was observed, showing deactivation in the dorsolateral and activation in the ventromedial parts simultaneously, which seems typical for any type of artistic, creative process (Fig. 7; see also [40] and text below): Dorsolateral deactivation stands for defocused attention combined with a lack of self-monitoring and volitional control, whereas the ventromedial activation may be interpreted as indicating basic attitudes and the preservation of rules. In addition, activity in the premotor and primary motor cortices indicated motor control as well as aspects of motor preparation (for playing a keyboard with the right hand in the scanner tube), whereas fine-grained adjustments of finger movements were regulated via cerebellar activity. Furthermore, emotionally inhibiting factors such as anxiety and fear seem to be suppressed during improvisation as the respective limbic areas, especially the amygdala, showed deactivation. Limb and Braun [28] conducted their study with six professional male pianists, each highly proficient in improvisation. While lying in the scanner, these pianists had to spontaneously modify a simple overlearned jazz melody on a small keyboard placed on their lap while listening to prerecorded combo chords via earphones.

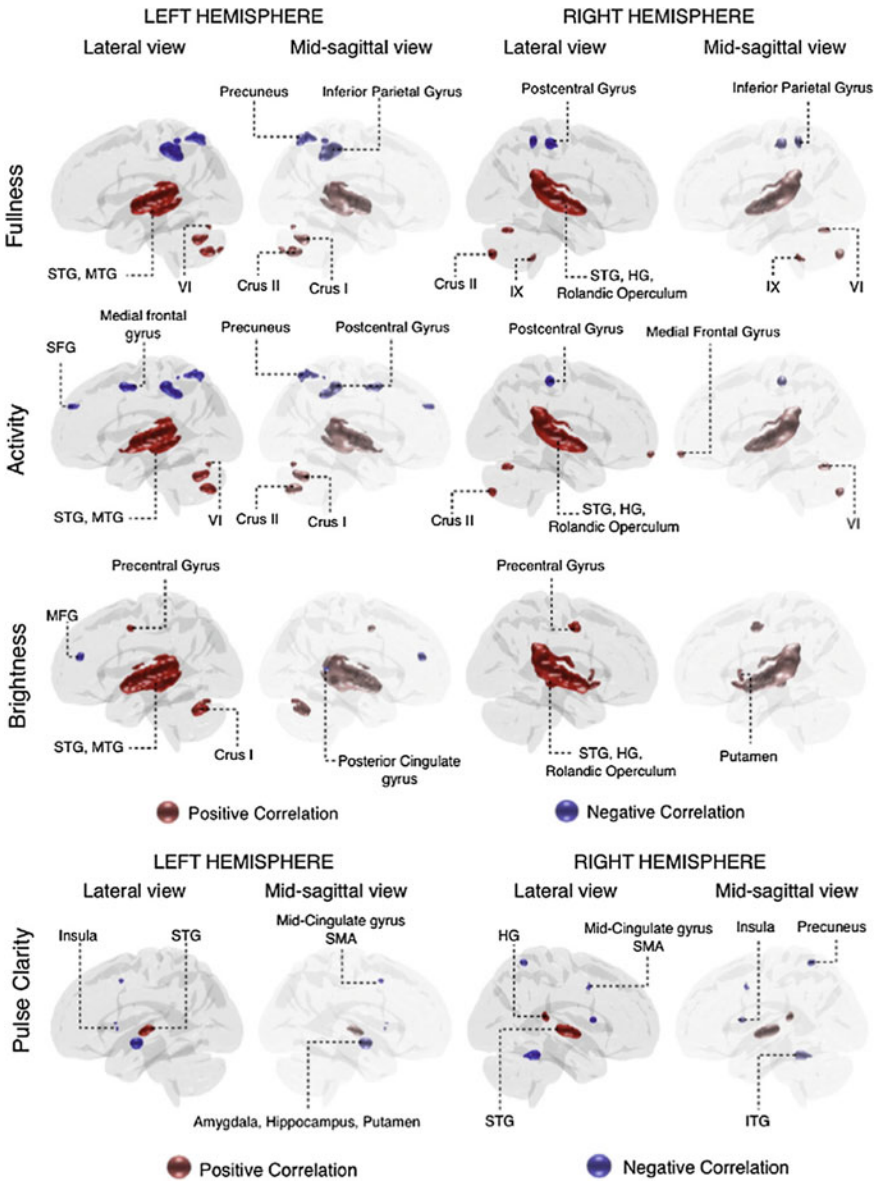


Fig. 6 Brain activity while listening to a complex piece of music—a modern tango. The auditory cortices are activated by high levels of pulse clarity and when processing several timbre features. Deactivations could be observed in the insula and the limbic system (for pulse clarity) as well as in parietal regions (for timbre attributes) [1]

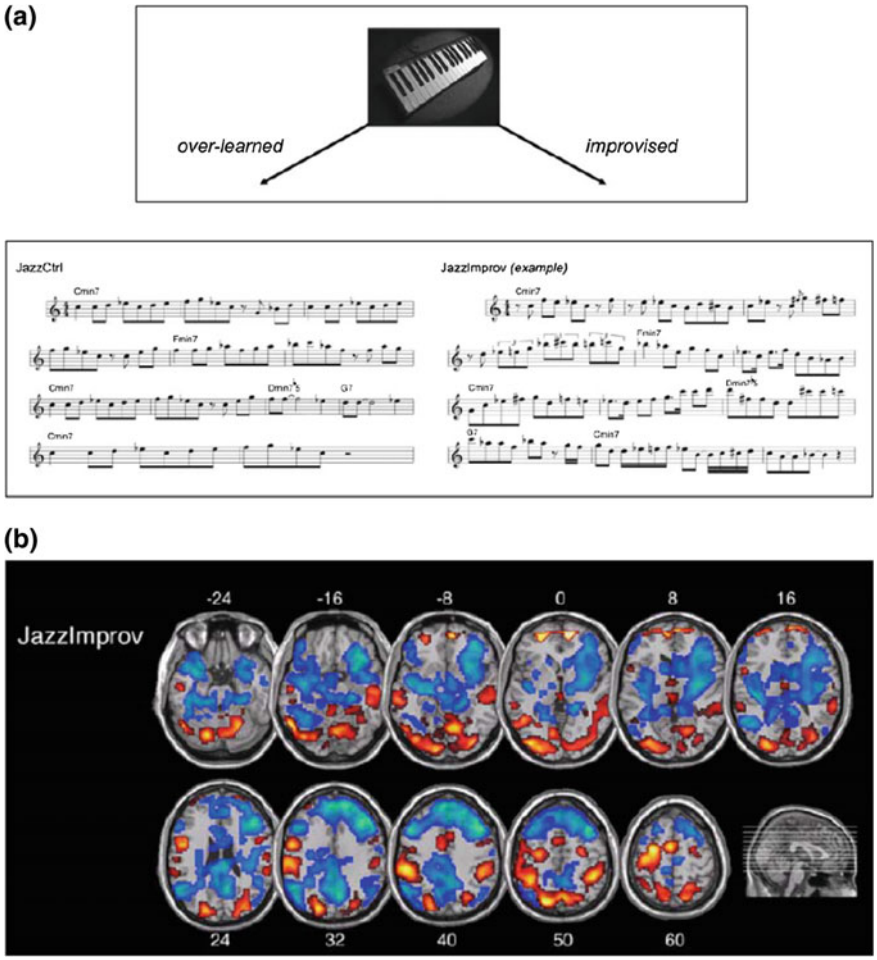


Fig. 7 **a**, **b** Activity in fMRI scans averaged over six male professional pianists. For any type of artistic creativity (here: jazz improvisation) a simultaneous division of the prefrontal cortex into deactivated (dorsolateral) and activated (ventromedial) parts can be observed [28]

2.6 Neuroplasticity in Musicians—Structural and Functional Types

Note that two decades ago, K. Anders Ericsson, an American psychologist of Swedish descent, developed a concept named 'deliberate practice', saying that high levels of proficiency (or expert performance) need years of intensive training especially in young adulthood (see [8]). Interestingly, deliberate practice leaves

'traces in the brain' through re-shaping its areas, a process widely known as *neuroplasticity*.

Neuroscientists distinguish between two forms, a functional and a structural type of experience-driven neuroplasticity. In the first case cortical activation strength, i.e. the susceptibility of brain tissue, is modified (see below), whereas in the latter significant enlargements, caused by an increase of dendritic branching as well as by intensification of synaptic strength, can be observed on the macro-level. A special method called 'voxel-based morphometry' enables researchers to precisely assess the extent of experience-induced anatomical changes while analyzing scans of the T1-weighed type (see [13] for details).

Münte et al. [32] considered the brains of professional musicians as the best fitting type to investigate these plastic changes, but also brains of sportspersons or chess players could serve as ideal models. Note that the brain's structural (and functional) changes seem to correlate with the age of learning to play a musical instrument in that effects are stronger the earlier piano or violin lessons start (i.e. typically before the age of 10) (see e.g. [7, 37, 38]).

A first result became visible in experienced string players, revealing an asymmetric (structural and functional) enhancement of the primary somatosensory cortex. The effect could clearly be detected for the fingering hand, i.e. for left hand fingers 2–5, but neither for the bow hand (right hand) nor for the left hand of a nonmusician control group [7].

Pantev et al. [37, 38] confirmed the fact of pure *functional neuroplasticity* in musical contexts: While listening inattentively to either piano versus pure tones (or to tones of familiar versus unfamiliar instrumental timbre) functional activity (or cortical dipole strength) was significantly enhanced in the auditory cortex of musicians. No similar effect could be observed for the nonmusician control group. Note that in these studies Pantev et al. [37, 38] decided for MEG (Magnetoencephalography) which is another non-invasive method of data recording enabling researchers to measure the brain's weak *electromagnetic* fields by using highly sensitive SQUID detectors. The distinguishing features of MEG are its high temporal as well as spatial resolutions, however, special software for source localization is required.

Another impressive result of neuroplasticity is beyond neuromusicology and refers to taxi drivers in London: Maguire et al. [30] could demonstrate that extensive experience in navigation, as typical for taxi drivers, causes a structural enlargement of the hippocampus, more specifically, a significant increase of gray matter density in the posterior hippocampal parts (Fig. 8).

3 Electroencephalography: The Basics

Let me move on to *electroencephalography (EEG)*, the oldest and most established type of neuroscience methods. By placing electrodes onto the head's surface, EEG research remains almost exclusively non-invasive. Note, however, that special

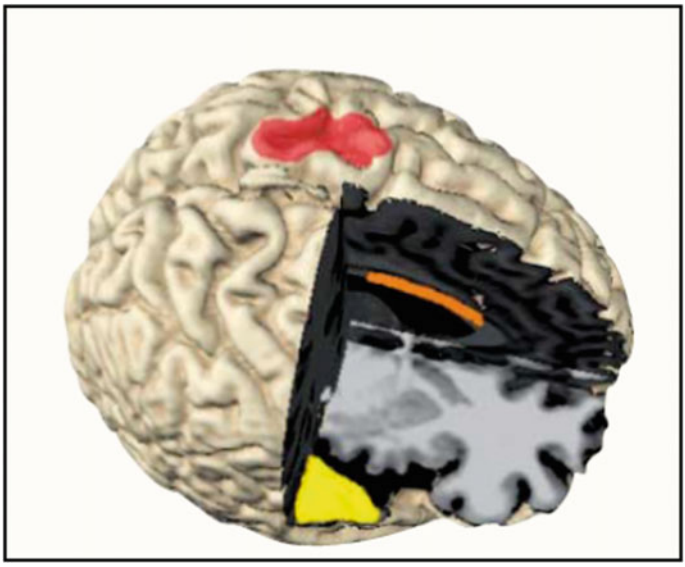


Fig. 8 The musician's brain serves as a model for neuroplasticity: structural enlargements of specific parts result from extensive training [32]

problems, for instance, localizing the source of musicogenic epileptic seizures during pre-surgical preparations, occasionally make it necessary, to either directly place electrodes onto the 'naked cortex', i.e. onto the gyri after opening the skull, or to implant them intracranially into the brain's tissue (e.g. [54]). The first variation is termed *electrocorticography (ECoG)*, the latter is called *depth EEG recording*.

The original, conventional EEG method was developed by Hans Berger, a German psychiatrist, at the beginning of the 1920s; he also coined the term 'electroencephalography' (Greek: enképhalon = brain; grapheîn = to write).

EEG potentials recorded from the scalp are raw signals, and each potential appears as the sum of extracellular field potentials stemming from different cortical layers. Extracellular field potentials, for their part, compensate intracellular voltage shifts when thousands of nerve cells 'fire' synchronously in order to solve a task. The electrogenesis of EEG scalp potentials can be described as follows (Tervaniemi and van Zuijen [55], p. 201):

> The EEG is a by-product of brain cells' information transfer in which intra- and extra cellular current flows are modulated with specific membrane mechanisms. When these current flows synchronize, potential differences summate, and become strong enough to be recorded with EEG. The post-synaptic activity of pyramidal dendrites (rather than action potentials) in the cortex particularly possess these characteristics and is therefore regarded as the main source of the EEG … Thus, in EEG, coherent activity of numerous cortical neurons (approximated by 10000) is recorded.

In most psychological contexts, EEG curves are registered by using *unipolar recordings* to measure potential differences between electrically active scalp

electrodes on the one hand and an electrically inactive reference point on the other (e.g., from an electrode placed on the earlobe). Furthermore, a standardized topographic schema named 'Ten-Twenty System', developed by a Canadian neuroscientist named Herbert Henri Jasper, helps to correctly place the scalp electrodes onto the head's surface. The 'Ten-Twenty System' creates a sort of individual anatomical coordinate system by taking two preauricular points as well as the nasion and the inion as reference points (see [17], for further details). This coordinate system enables researchers to precisely describe, compare or exchange sets with EEG data between research institutes (on the assumption that paradigms are equivalent) (Fig. 9).

Psychologically oriented EEG research often uses the Fast Fourier Transform algorithm (FFT) to separate four main frequency bands from each other (δ, θ, α, β; Fig. 10). Each band reliably shows a specific state of consciousness and/or level of arousal, ranging from 'coma', 'trance' and 'deep sleep' (indicated by a predominance of delta activity with oscillations between 0.5 and 4 Hz), to 'meditation' and 'drowsiness' (predominance of theta activity with oscillations between 4 and 8 Hz), followed by the state of 'being awake and relaxed' (predominance of alpha; 8–13 Hz) and the state of 'being mentally active and attentive' (predominance of beta, 13–30 Hz).

Note that a fifth type, the gamma band with frequencies between 30 and 80 Hz, is omitted in this context here: Gamma activity indicates 'feature binding', a specific process necessary to experience coherent percepts. It has mainly been found in the visual domain for binding spatially separate, 'static' properties together, like 'color', 'shape' and '(surface) texture'. Regarding the auditory domain, similar processes of temporal feature binding have been found less frequently. (Despite that [4], observed stronger gamma-band-synchronizations in musicians than in non-musicians when listening to musical excerpts, suggesting that musicians are more experienced in anticipating melodies of a certain style as well as in combining musical parameters such as contour, rhythm and timbre to a melodic entity.)

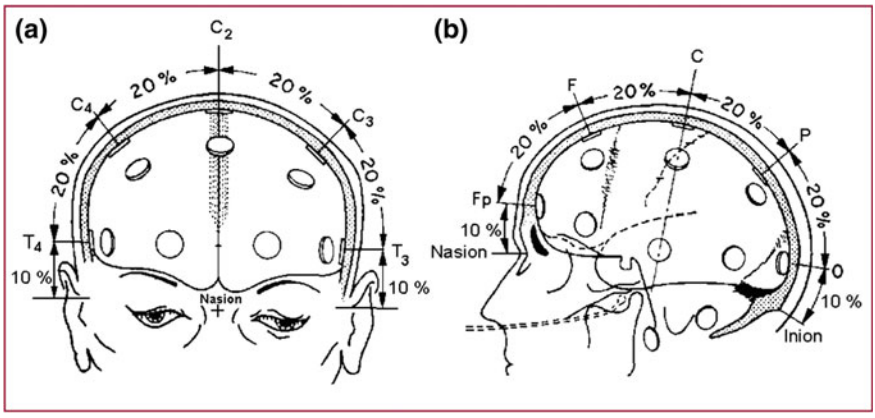


Fig. 9 Electrode placements according to the 10–20 system [17]

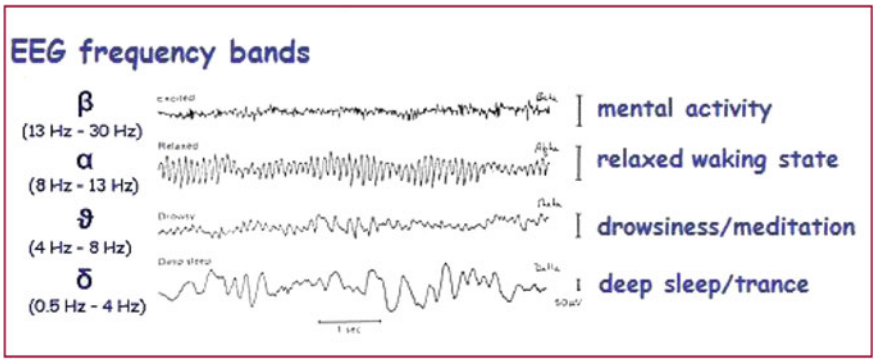


Fig. 10 Fourier analysis: the EEG raw signal is divided into four frequency bands. Each indicates a specific state of consciousness and arousal (gamma activity is missing)

3.1 Research with EEG: Two Example Studies

It is common knowledge among ethnologists that states of Shamanic trance can be reached by taking drugs and/or by repetitive, monotonous drumming (e.g. [11, 44]). The brain reacts to this mind-expanding experience with a change in the spectral content of the standard EEG, in particular by an increase of theta and delta activity.

To elicit a form of sound-induced trance in a Western context, Kohlmetz et al. [23] chose a special piece of piano music named 'Vexations' (1893), written by the 27-year-old French composer Erik Satie (Fig. 11). Satie suggested playing this, per se, short (atonal and contrapuntal) composition 840 times without interruption, resulting in a total performance length of approximately 28 h. Here, trance is induced by an unusual playing instruction. Armin Fuchs, a German pianist, succeeded in getting through this 28-h 'endurance test' while the EEG was simultaneously recorded from parietal electrode sites. Despite mental stress and physical strains Fuchs kept tempo and motor performance relatively stable. He reported having experienced a 5-h-state of trance, in addition to a feeling of slowing down and that of lengthened time [23]. During trance, brain activity decreased bilaterally towards the delta-band and shifted slightly towards the left parietal electrode (P3). Note that only two electrodes were placed in the upper back of the pianist's head to avoid motion restrictions. However, from the viewpoint of EEG recording, this is below the minimum required by guidelines to fulfill the criterion of reliability.

Let me briefly describe a second example of music-related EEG research: During the 1980s, the Austrian neuroscientist Hellmuth Petsche came up with the idea of 'EEG coherence analysis', a methodological approach performed for each frequency band (δ, θ, α, β) separately to extend the conventional type of analyzing EEG raw signals via FFT. EEG coherence analysis has proven highly effective to investigate the interplay between cortical network structures during creative thinking and other mental processes of higher order. It describes the degree of similarity (or functional coupling) between EEG signals at adjacent electrodes of

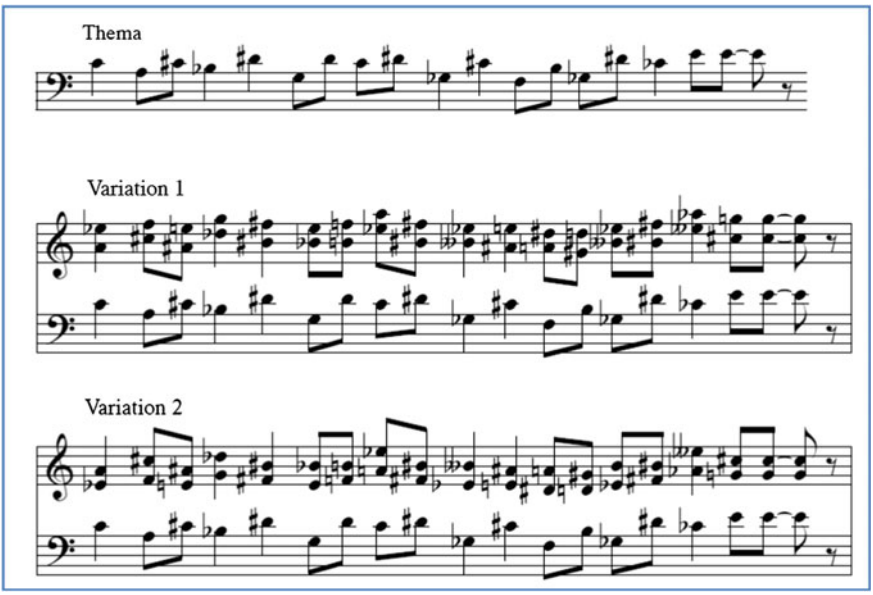


Fig. 11 'Vexation', a piano piece by Erik Satie: It has to be played 840 times without interruption, resulting in a performance duration of 28 h and the experience of trance

the same hemisphere (the 'intrahemispheric type') or at homologous electrode sites on the opposite halves of the brain (the 'interhemispheric type').

In an EEG coherence study on composition, Petsche [40] put this approach into practice: Seven male professional composers had the task to spontaneously invent a tonal passacaglia and an atonal fugue, each of 5 min length, and write these pieces down immediately afterwards. Figure 12 shows the coherence patterns for one 56-year-old male composer while listening to a piece by Schönberg (the control condition) and while mentally composing in both styles. The most striking result appears as a shift of activation from left inferior-frontal regions (Fig. 12, left side), reflecting syntax analysis, towards the posterior parietal cortex (Fig. 12, middle), probably indicating some thoughts about the formal shape, or 'musical architecture' of the piece in progress.

Besides that, the study revealed two more results: First, any type of creative process in art (be it verbal, visual or musical) is indicated by a functional decoupling (or decrease of coherence) in the dorsolateral prefrontal cortex (dlPFC) so that bizarre, uncontrolled thoughts can enter (see also [5] and [28]). Second, long-distance coherences, for instance between fronto- and parietal electrode sites, increase during the mental act of composing, and interindividual differences are high (Fig. 13).

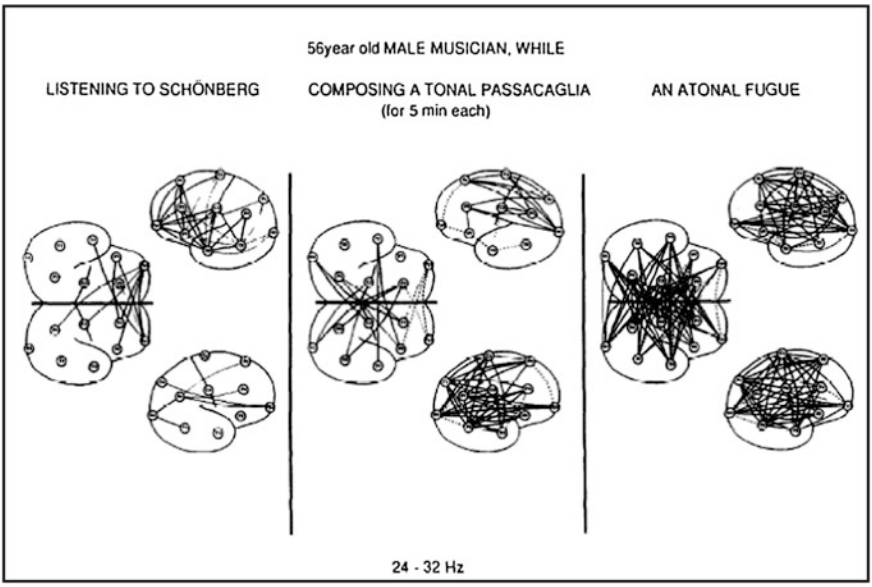


Fig. 12 Result of an EEG coherence analysis (beta band) for a 56-year old male composer: *Left* listening to Schönberg, *middle* composing a tonal passacaglia, *right* composing an atonal fugue (three views of the brain: from *above*, *left hemisphere*, *right hemisphere*) [40]

3.2 EEG Sports: A Promising Trend Using Mobile Devices

Investigating brain activity of humans *in action*, while playing golf, riding a bicycle or performing in a chamber music ensemble has been an unsolved problem in EEG research for many years. The most challenging aspect is not the real life situation per se but rather body movement as such: Any subject in motion produces many artifacts of extracerebral origin, arising from skin changes and muscle tension. In addition, sweating and loose electrodes may also contaminate the measurement and make EEG data not utilizable for further analysis (e.g. [56]). Furthermore, the recording equipment is unwieldy and heavy, including amplifiers and batteries, making it impossible to carry the device in a rucksack on the back. On the other hand, portable solutions would offer a wealth of opportunities in the field of human movement and sports science, while ecological validity would be high.

Until now, most studies in this context use EEG for neurofeedback-training in the lab, i.e. for investigating brain-based self-regulatory techniques that may help to modify the mental attitudes during several phases of practicing and performance. This way, certain EEG frequency bands, in particular alpha, theta and delta, can reliably be strengthened via monitor and other feedback devices, obviously increasing self-awareness, feelings of well-being and the supply of mental and physical energy necessary to succeed in any training session or sports competition outside [39, 56].

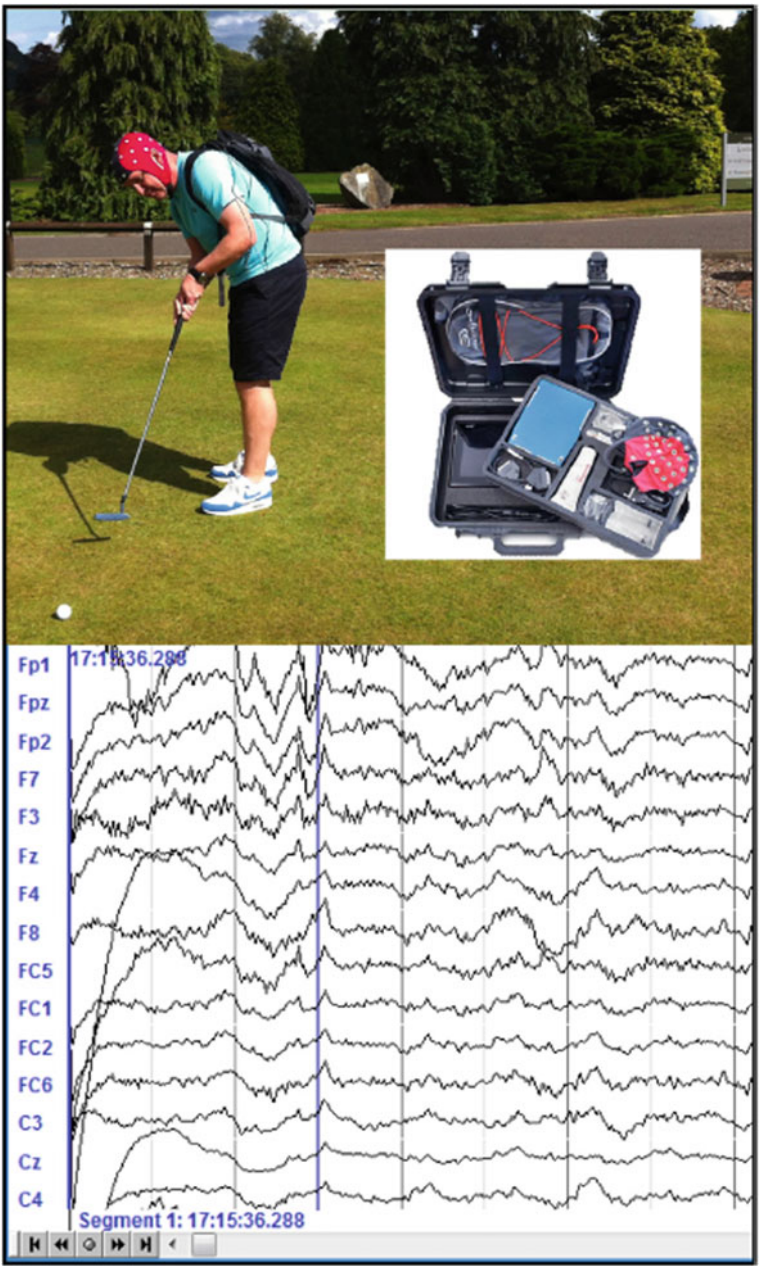


Fig. 13 A newly developed portable EEG solution, termed **eego**sports™, enables researchers to investigate brain activity in real life situations [39]

Recently, a new product series, termed **eego**sports ™ has been developed by ANT Neuro, Enschede, a neuroscience company specialized in developing EEG hard- and software. Since 2013, they offer a portable, light-weight 64 channel EEG solution of less than 1000 g that enables researchers to freely investigate different types of movement as well as effects of training and physical exercise in a natural environment. Presumably, this mobile solution will be used in the context of 'music and motion' in the near future.

4 Event-Related Potentials (ERPs)—A Derivative of the EEG

Finally, let me describe the second type of bioelectric methods, known as the measuring of *'event-related potentials' (ERPs).* ERP works on the precondition that, during recording, the same type of stimulus will be repeated at least 20 times which is not necessarily required for measuring the EEG.

Both methods, EEG and ERP, differ completely in their basic idea: EEG, on the one hand, allows to make individual recordings of several minutes length while disregarding transient brain activity, i.e. the components lasting some ms within ultra-short time frames. This way, the EEG informs about the brain's overall physiological state, i.e. the levels of consciousness and arousal while listening to music of different style and tempo.

ERP, on the other hand, is completely devoted to the basic idea of drawing an analogy between the computer and the human mind, meaning that both systems, the electronic and the human, should be considered similar in their strategies to select, transform, store or activate the respective information (see [34]). The ERP, therefore, directly points to a, mainly, serial form of processing input and comprises several independent processing steps in sequence [13]. (Note that according to this shift in thinking the word 'cognition', derived from the Latin word 'cognoscere', has lost its former philosophical connotations like 'becoming aware of', 'apperceive' or 'comprehend intuitively' and is now used in a simple, pragmatic way).

How are ERP results obtained? First of all, ERP uses the same initial recording procedure as EEG, so that raw traces as such cannot be classified as either belonging to the first or second type of method. Because of this, a special form of data analysis, termed *signal averaging*, is required to extract the amount of event-related brain activity from the raw data: Signal averaging enables researchers to split the raw signal into a spontaneous and a stimulus-related part by taking advantage of the fact that each stimulus repetition evokes a small but invariant brain response that can be summed up and divided by the number of presented stimuli, whereas spontaneous, stimulus-*uncorrelated* fluctuations converge against zero (Fig. 14). To further increase the signal-to-noise ratio, individual ERP curves are

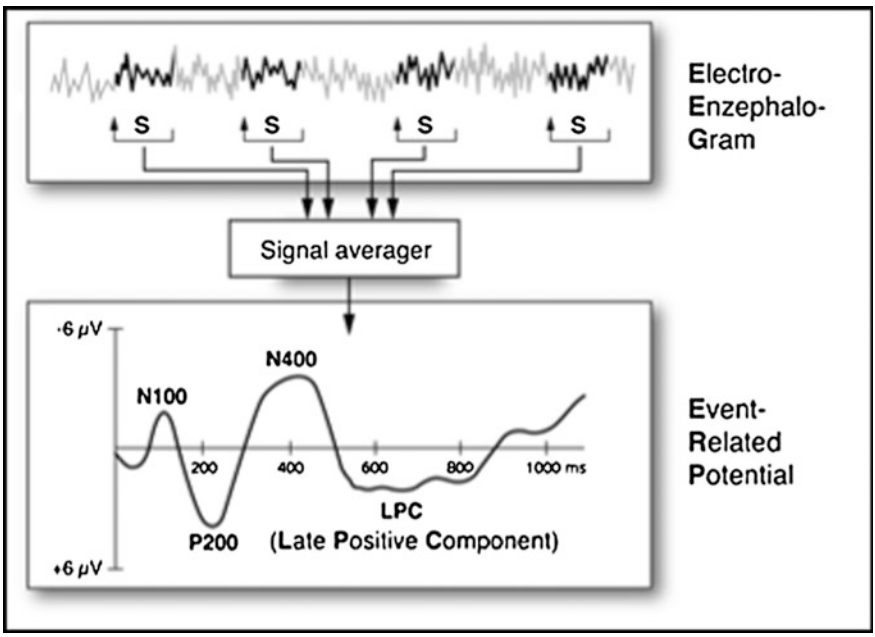


Fig. 14 Principle of signal averaging: marked EEG sections contain brain responses assigned to a specific stimulus S. They are summed up and divided by the number of stimulus repetitions, yielding the event-related potential. It informs about various steps of processing incoming information (MPI Cognitive and Brain Science Leipzig)

finally summed up to group-wise potentials, the so-called *grand average ERPs*, and this is the starting point for the analysis of ERP components (see below).

Rösler [43] points out that signal averaging and its product, the grand average ERP, are flawed with some weak points: First, brain responses stemming from several trial repetitions are summed up automatically which is considered inappropriate from a psychological point of view as it cannot be ruled out that participants might have changed attentiveness during recording. Second, brain responses are prone to habituation, i.e. amplitudes will be reduced the more familiar, or predictable, the often-repeated stimuli are. Third, grand average ERPs are produced at the expense of individual brain responses, meaning that conclusions regarding individual processing strategies cannot be drawn from the final product.

Note that it is not the ERP curve as a whole that serves as a unit for interpretation. Instead, each half wave, or ERP component, will be analyzed separately on the assumption that it responds independently, i.e. without any cohesive forces operating between adjacent components.

Regarding nomenclature, two details are needed to describe each ERP component properly: details about its 'polarity' and its 'latency'. The term 'polarity' describes the component's deflection, i.e. change in voltage direction either into the positive ('P') or the negative ('N'). 'Latency', by contrast, refers to the timespan

between stimulus onset and the peak amplitude and can either be described as a rounded value (in ms; e.g. P200) or as an ordinal number (e.g. P2). These sparse but essential details may be completed by some more information about the component's 'maximum amplitude value' [μV], its 'brain topography' and its 'waveshape'.

The curve example in Fig. 14 shows five components, termed P50, N100, P200, N400 and LPC (late positive component). The first ones up to 300 ms indicate exogenous processes that, in principle, are determined by stimulus parameters such as frequency, intensity or presentation rate. N400 and LPC, by contrast, indicate endogenous processes, reflecting some task-related cognitive processing steps for which attention is required. However, since recent results could show that exogenous components can be modulated by top-down processes too (e.g. [33]), contemporary ERP research directly focuses on the characteristics of the particular component itself, i.e. it omits this additional exogenous-endogenous classification.

To illustrate which aspects of cognitive processing a component may indicate, I will first pick out the so-called Mismatch Negativity (MMN).

4.1 The 'Mismatch Negativity' (MMN)—An Example Component of the ERP

Imagine you hear the following example: AAAA AAAA AA**B**A AAAA A**B**AA AAAA, i.e. a basic type of sequence in which two elements, a frequently repeated pure tone A (the standard) and an occasionally inserted pure tone B (the deviant) form an auditory chain which can be partitioned into bunches of four (the ratio of A to B may be 0.9:0.1 or 0.75:0.25). Interestingly, the brain will not only process the acoustic parameters of A and B separately (by developing a P50 and an N100), but will also react to this irregular and unpredictable alternation of items within the sequence. The detection of deviant 'B' will be indicated by a specific ERP component, called Mismatch Negativity (MMN) on the implicit assumption that a (temporary) memory trace has already been built for the regular sequence AAAA AAAA (Fig. 15). The MMN has its origins in the primary and secondary auditory cortices and the right frontal lobe, thus, indicating detection processes mainly for the auditory domain. Its maximum amplitude is between 130 and 250 ms measured from deviant onset.

From a functional point of view, the MMN indicates some trace-building processes within sensory (echoic) memory which gives rise to the assumption that a pattern-based process is the underlying driving force [20]. The second attribute is its independency from attention, enabling researchers to investigate both, attentive (controllable) as well as preattentive (automatic) processes of sound detection. Regarding the latter, attention is caught by instructing subjects to watch a silent video or read a book which prevents them from taking particular notice of the sounds themselves.

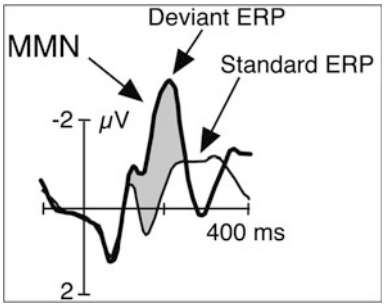


Fig. 15 The MMN is the first component that does not react to the properties of a single tone, but rather to some irregularities within the auditory sequence. Subjects may either focus on the tone series as such (attentive condition) or be distracted by watching a movie (preattentive condition). Adapted from Kujala and Näätänen [24]

Interestingly, these preattentive mechanisms of sound detection are modifiable by longstanding experience in that amplitudes are higher the more musical training participants have, in other words, the more accurate sound-templates in long-term memory are stored: Violinists, who are well-experienced in shading intervals and chords according to good intonation, automatically detect a slightly impure major chord (with frequencies of 396-491.25-596 Hz instead of 396-495-596 Hz), and this discrepancy between the actual input and the stored template will be indicated by a clear MMN. Musically inexperienced participants do not show a similar result [21].

4.2 Syntactic and Semantic Incongruities in Language and Music: ELAN/ERAN, P600 and N400

As already seen in the previous paragraph, ERP works best when tone rows are investigated, that is, when structure unfolds along the time axis. This way, sequence structures of any type will match the method's distinguishing feature of registering transient brain activity in high resolution on a ms-time scale.

By using a different paradigm, three specific ERP components have been found in the language domain, named ELAN (early left anterior negativity), N400 and P600, that are connected with a *rule-based type of sequence structure*: the ELAN and P600 indicate error detection in terms of syntax structure, whereas the N400 indicates deviation regarding semantic content.

In more detail, omissions with regard to word category ('the blouse was on [ ] ironed') elicit an ELAN between 150 and 200 ms (measured from onset of the word 'on'), whereas the late P600 indicates some sort of conscious re-analysis of the entire sentence structure (Fig. 16). Thus, ELAN and P600 are two ERP components

reflecting both, the preconscious (automatic) as well as the controlled aspects during the initial and later phases of syntax analysis [10].

Interesting parallels can be drawn between syntax processing in language and music: That is, melodies ending with a non-diatonic, incongruous final tone [3] evoke early and late components (named ERAN [early right anterior negativity] and P600) of similar shape and latency as those ones that were previously found for processing spoken sentences, allowing the conclusion that underlying processes are domain-general (Fig. 16).

(Note that similar comparative results were achieved when processing prosodic information: A specific ERP component termed Closure Positive Shift (CPS) was found for processing intonational phrase boundaries in spoken language as well as for processing phrase boundaries in music (while listening to binary-structured melodies); cf. [18, 35, 51]).

Besides that, a prominent component, termed N400 reacts to challenges of the semantic type: The N400 is visible whenever absurd or meaningless words in otherwise grammatically correct sentences are identified ("He carries his daughter in his *nostrils*"). The N400 is therefore interpreted as indicating violations of semantic expectancy [25].

It is worth investigating how the N400 'behaves' in the context of music, since music is commonly regarded as being more ambiguous than language. Koelsch et al. [22] approached this question by creating a priming situation, i.e. by presenting either a spoken sentence or a musical excerpt *preceding* a word, serving as a target stimulus. An N400 was evoked after both types of priming, making Koelsch suggest that a common associative network might be activated (Fig. 17).

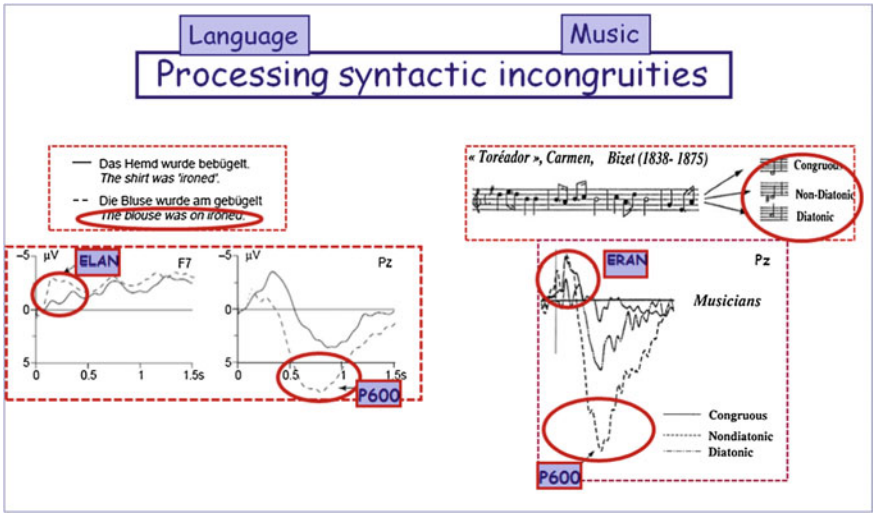


Fig. 16 Syntactic incongruities in spoken sentences (*left*) and in melodies (*right*) elicit similar components, termed ELAN/ERAN and P600 [3, 10]

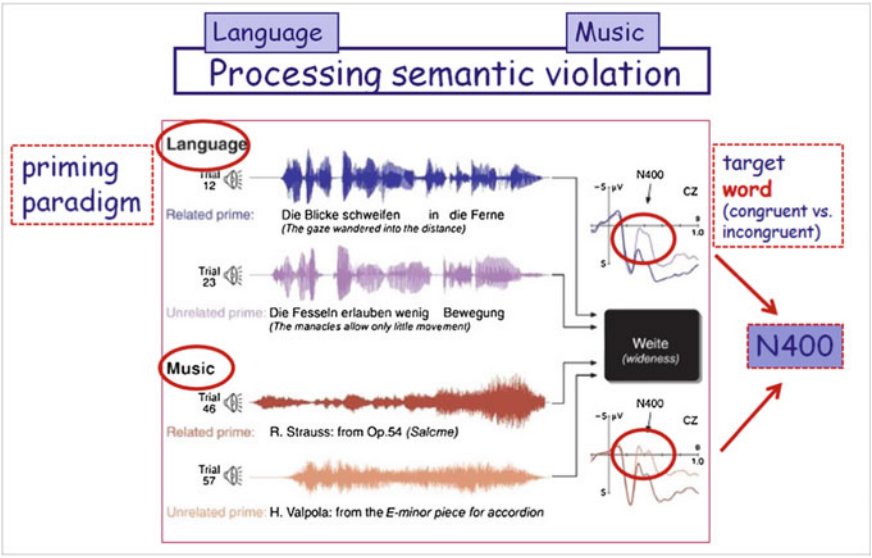


Fig. 17 An incongruous target word ("Weite") elicits an N400. It is unrelated to the previous context taken from music or the language domain [22]

To my knowledge, no study exists at present where the N400 is evoked by a semantic mismatch between a *musical context* on the one hand and a *musical target* (e.g. a chord) on the other, supporting the widespread view that chords, intervals and musical excerpts are less clear in meaning than words.

5 Do Advantages Outweigh the Disadvantages?—A Final Assessment of the Methods' Pros and Cons

Let me sum up the latest developments in neuromusicology. In my view, the following three tendencies become apparent: First, there is the endeavor to precisely relate *cause and effect*, i.e. to prefer causal relationships to correlative ones. This means that transcranial magnetic stimulation (TMS) is increasingly applied to music-related questions [26, 52], allowing researchers to assign an either slowed down or accelerated overt musical behavior to differently stimulated brain tissue.

The second trend is towards investigating brain activity in *real life situations*, i.e. to increasingly fulfill the criterion of *ecological validity*. Recently developed EEG mobile solutions (**eego**sports^{TM}) match well with this concept: They enable researchers to investigate the 'brain in action' while sledging, jogging, preparing a solo recital or playing in a jazz combo, in short: during every type of sport and performance activity. This trend also includes the endeavor to record brain activity

in natural environments, for instance, while performing cross-cultural field studies in non-Western countries.

Third, several labs advocate a holistic approach in that *whole-brain activity* is explored with fMRI while listening to complex musical pieces or while spontaneously improvising on the piano [1, 28]. Regarding this holistic approach, EEG coherence analysis, developed by Petsche as early as 1996, might be considered as a forerunner, since functional coupling of cortical network structures (while composing or listening to short pieces of music) can also be investigated by using this type of analysis.

Starting the assessment of methods with *EEG*—What are the main advantages and disadvantages of this oldest and most established type of neuroscience methods?

EEG allows the recording of *unspecific brain activity* over a time span of several minutes length while disregarding the transient components. Accordingly, the EEG is an appropriate method whenever *experiencing music*, be it repetitive drumming or a Mozart symphony. So, EEG is the method of choice whenever the focus of research is on the level of consciousness, on attention or arousal. Furthermore, Fourier analysis enables researchers to precisely observe changes in the spectral content of EEG signals over the entire time of recording. A disadvantage might be that interpretation is limited to statements about the brain's physiological state in general. However, a further plus point is coherence analysis, an option showing the functional coupling of brain activity at near and distant electrode placements, yielding information about the interplay between cortical network structures. So, whenever coherence analysis is included as an additional tool, the method's power is considerably increasing.

A second advantage is the possibility to analyze EEG raw traces for each subject separately. This is without alternative whenever mental states during creative processes are investigated, making the EEG method indispensable for creativity research (Schaffenspsychologie). However, there is also an option for a group-wise EEG analysis by comparing frequency bands of, e.g. musicians vs. non-musicians in relation to a specific task or a particular piece of music.

What are the pros and cons of measuring *'event-related potentials' (ERPs)*, the second type of bioelectric methods?

In principle, ERP components indicate information processing in a step-by-step manner. This way, ERP supports the basic idea of cognitive psychology which says that the human mind and the computer work on analogous principles [34]. In addition, excellent time resolution allows the discovery of new components beyond the established ones, for instance the face-sensitive N170, indicating the earliest stage of face recognition (e.g. [46]).

Regarding contents, ERP is the appropriate method to investigate three specific aspects: (a) to examine brain responses to frequency, intensity and other sound parameters in the context of psychoacoustics (eliciting a P50, an N100 and a P200, respectively), (b) to investigate brain responses to (simple) auditory sequence structures (eliciting an MMN) and (c) to examinethe processing of rule-based

syntax structures (evoking components such as ELAN/ERAN and the P600) so that comparisons between language and music can be drawn.

However, a lot of methodological constraints are imposed on each ERP design with restrictive effects on the interpretation of results: First, to increase the signal-to-noise ratio—between the event-related potential and spontaneous, unrelated EEG activity—the same type of stimulus has to be repeated between 20 and 1000 times (depending on the respective paradigm). Second, grand average ERPs refer to the entire number of trials and subjects, making it impossible to measure inter-individual differences or to reconstruct individual responses in retrospect. Third, brain responses are prone to habituation which means that, for each subject, trials are subsumed to a single common curve while disregarding minor or even major shifts in attention which is considered inappropriate from a psychological point of view.

Fourth, empirically working musicologists should know that ERP does not allow any conclusion about processing musical pieces of a particular epoch, a specific genre or the personal (idiosyncratic) style of composer (e.g. Händel vs. Bach). Fifth, the ERP responds to structure-violation in an overall-sense, for instance to any deviant chord within the standard scheme, be it a Neapolitan Sixth (N^6 or s^n) or a double dominant (D^D) in a musical cadence. These types of deviant chords will evoke the same ERP component (ERAN), no further specification in terms of harmonic progression will be possible.

Finally, let me weigh up the main advantages and disadvantages of *fMRI* and *PET*, the most popular types of neuroimaging methods:

Due to an excellent resolution in the spatial domain (approximately 1–4 mm^3 for fMRI and 2–8 mm^3 for PET depending on the scanner type) both neuroimaging methods provide the possibility to localize even the smallest functionally activated brain areas, based on a voxel-wise analysis. This way, the complex interplay between cortical and subcortical network structures, including the basal ganglia, the cerebellum and parts of the brain stem, can be made visible.

However, the precise 'Where' in the brain is at the expense of the 'When': The BOLD signal reaches its peak plateau between 4 and 8 s after task-onset, thus, in comparison to EEG and ERP, time resolution is poor.

Among the various options of stimulating brain tissue with pulse trains and sharp HF-impulses to obtain high-quality imaging data, echoplanar imaging is the fastest, enabling researchers to record the whole brain in less than 2 s. However, a disadvantage of this stimulation type is the technical noise in the scanner with volume intensities between 60 and 100 dB due to a fast switching of gradient coils during space-coding, a necessary step for image acquisition. Note that the interleaved silent steady state method is the most sensitive of all echo-planar imaging techniques, suitable for detecting subtle activities in subcortical structures as well as in deeper layers of the cerebrum [31].

However, in terms of musically-related contexts neuroscientists often choose PET, the older one of both imaging techniques. PET in contrast to fMRI does not produce any disturbing scanner noise at all. This enables participants to deepen their emotional experience and perceive a musical piece in an aesthetic sense while lying

in the scanner. However, the major disadvantage of the PET method is that a radioactive tracer substance has to be intravenously injected. This procedure imposes several restrictions on the recording procedure and the choice of stimuli, both aspects are strictly determined by the rate of decay.

To sum up: Do advantages outweigh the disadvantages? The question should be answered in the positive: Neuroscience methods offer elegant solutions to measure cognitive processes in real-time, yielding results of either high-temporal or high-spatial resolution. This fits nicely with a proposal by Leman [27]: To solve research problems more successfully he recommends a "joint correlative approach between different research methodologies; in particular musicology, computer modeling, experimental psychology and […] neuromusicology" (p. 194f), in short, a "convergence paradigm" . This is in accordance with a truly systematic approach as advocated by Schneider [47]: "The ultimate goal of systematization is to establish coherent systems of knowledge that should be free from any contradictions, and should be as complete in descriptions and explanations of the principles and phenomena of a certain field as is possible." (p. 20).

However, to dampen euphoria and overoptimism regarding the available neuroscience methods and their capacities, take notice of the following:

"The goal of neural science is to understand the mind—how we perceive, move, think, and remember." Despite all efforts, this statement by Eric Kandel (cited in [58], p. 1) still cannot be put into practice. (Some experiments on mental rotation make an exception, e.g. [15, 48]). Until now, many impressive methods inform about the *physiological state* and the *functional activity of the brain*. But how the mind works, is a different matter. Scientists still do not know for sure how thoughts are generated and how mental knowledge representations precisely look like. Nevertheless, attempts have recently been made to reconstruct the mental content belonging to different semantic categories from fMRI scans showing cortical activation and deactivation (e.g. [12]). However, the main disadvantage of this approach is ambiguity in that, until now, no clear assignment between both types of substance, the material and the immaterial world, can be made. Even so, innovations in the field of neuroscience are growing rapidly, so there are grounds to believe that the dualism between mind and brain, the so-called 'hard problem', may be solved in the near future.

References

1. Alluri, V., Toiviainen, P., Jääskeläinen, I.P., Glerean, E., Sams, M., Brattico, E.: Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm. NeuroImage **59**, 3677–3689 (2012)
2. Andoh, J., Zatorre, R.J.: Interhemispheric connectivity influences the degree of modulation of TMS-induced effects during auditory processing. Front. Psychol. **2**, Article 161, 13 pages (2011). doi:10.3389/fpsyg.2011.00161

3. Besson, M., Faïta, F.: An event-related potential (ERP) study of musical expectancy: comparison of musicians with nonmusicians. J. Exp. Psychol.: Hum. Percept. Perf. **21**(6), 1278–1296 (1995)
4. Bhattacharya, J., Petsche, H., Pereda, E.: Long-range synchrony in the γ-band: role in music perception. J. Neurosci. **21**(6), 6329–6337 (2001)
5. Dietrich, A.: The cognitive neuroscience of creativity. Psychon. Bull. Rev. **11**, 1011–1026 (2004)
6. Drobisch, M. W.: Über musikalische Tonbestimmung und Temperatur [On musical pitch estimation and temperature]. In: Abhandlungen der Königlich-Sächsischen Gesellschaft der Wissenschaften **2**, 1–120. Hirzel, Leizpig (1855).
7. Elbert, T., Pantev, C., Wienbruch, C., Rockstroh, B., Taub, E.: Increased cortical representation of the fingers of the left hand in string players. Science **270**(5234), 305–307 (1995)
8. Ericsson, K.A.: The influence of experience and deliberate practice on the development of superior expert performance. In: Ericsson, K.A., et al. (eds.) The Cambridge Handbook of Expertise and Expert Performance (Chapter 38, pp. 685–706. Cambridge University Press, New York (2006)
9. Fachner, J.: Topographic EEG changes accompanying Cannabis-induced alteration of music perception—Cannabis as a hearing aid? J. Cannabis Ther. **2**(2), 3–36 (2002)
10. Friederici, A.D.: Towards a neural basis of auditory sentence processing. Trends Cogn. Sci. **6** (2), 78–84 (2002)
11. Gingras, B., Pohler, G., Fitch, W.T.: Exploring Shamanic journeying: Repetitive drumming with Shamanic instructions induces specific subjective experiences but no larger Cortisol decrease than instrumental meditation music. PLOS One **9**(7), 9 pages (2014)
12. Haynes, J.-D., Rees, G.: Decoding mental states from brain activity in humans. Nat. Rev. Neurosci. **7**, 523–534 (2006)
13. Jäncke, L.: Methoden der Bildgebung in der Psychologie und den kognitiven Neurowissenschaften. W. Kohlhammer, Stuttgart (2005)
14. Jäncke, L.: Lehrbuch Kognitive Neurowissenschaften. Huber, Bern (2013)
15. Jäncke, L., Jordan, K.: Functional neuroanatomy of mental rotation performance. In: Mast, F. W., Jäncke, L. (eds.) Spatial Processing in Navigation, Imagery and Perception, pp. 183–207. Springer, New York (2007)
16. Janata, P., Birk, J.L., van Horn, J.D., Leman, M., Tillmann, B., Bharucha, J.J.: The cortical topography of tonal structures underlying Western music. Science **298**, 2167–2170 (2002)
17. Jasper, H.H.: The ten-twenty electrode system of the international federation. Electroencephalogr. Clin. Neurophysiol. **10**(2), 370–375 (1958)
18. Knösche, T.R., Neuhaus, C., Haueisen, J., Alter, K., Maess, B., Witte, O.W., Friederici, A.D.: Perception of phrase structure in music. Hum. Brain Mapp. **24**(4), 259–273 (2005)
19. Köchli, V.D., Marincek, B.: Wie funktioniert MRI?. Springer, Berlin (1998)
20. Koelsch, S.: Music-syntactic processing and auditory memory: similarities and differences between ERAN and MMN. Psychophysiology **46**, 179–190 (2009)
21. Koelsch, S., Schröger, E., Tervaniemi, M.: Superior pre-attentive auditory processing in musicians. NeuroReport **10**, 1309–1313 (1999)
22. Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., Friederici, A.D.: Music, language and meaning: brain signatures of semantic processing. Nat. Neurosci. **7**, 302–307 (2004)
23. Kohlmetz, C., Kopiez, R., Altenmüller, E.: Stability of motor programs during a state of meditation: Electrocortical activity in a pianist playing 'Vexations' by Erik Satie continuously for 28 hours. Psychol. Music **31**(2), 173–186 (2003)
24. Kujala, T., Näätänen, R.: The mismatch negativity in evaluating cental auditory dysfunction in dyslexia. Neurosci. Biobehav. Rev. **25**(6), 535–543 (2001)
25. Kutas, M., Hillyard, S.A.: Reading senseless sentences: brain potentials reflect semantic incongruity. Science **207**, 203–208 (1980)

26. Launay, J., Dean, R.T., Bailes, F.: Rapid learning of associations between sound and action through observed movement. A TMS study. Psychomusicology **26**(1), 35–42 (2016)
27. Leman, M.: Relevance of neuromusicology for music research. J. New Music Res. **28**(3), 186–199 (1999)
28. Limb, C.J., Braun, A.R.: Neural substrates of spontaneous musical performance: an fMRI study of Jazz improvisation. PLoS One **3**(2), e1679 (11 pages) (2008)
29. Logothetis, N.K.: What we can do and what we cannot do with fMRI. Nature **453**, 869–878 (2008)
30. Maguire, E.A., Gadian, D.G., Johnsrude, I.S., Good, C.D., Ashburner, J., Frackowiak, R.S.J., Frith, C.D.: Navigation-related structural change in the hippocampi of taxi drivers. PNAS **98** (8), 4398–4403 (2000)
31. Mueller, K., Mildner, T., Fritz, T., Lepsien, J., Schwarzbauer, C., Schroeter, M.L., Möller, H. E.: Investigating brain response to music: a comparison of different fMRI acquisition schemes. NeuroImage **54**, 337–343 (2011)
32. Münte, T.F., Altenmüller, E., Jäncke, L.: The musician's brain as a model of neuroplasticity. Nat. Rev. Neurosci. **3**, 473–478 (2002)
33. Musacchia, G., Sams, M., Skoe, E., Kraus, N.: Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. PNAS **104**(40), 15894–15898 (2007)
34. Neisser, U.: Cognitive Psychology. Meredith, New York (1967)
35. Neuhaus, C., Knösche, T.R., Friederici, A.D.: Effects of musical expertise and boundary markers on phrase perception in music. J. Cogn. Neurosci. **18**(3), 472–493 (2006)
36. Ogawa, S., Lee, T.M., Kay, A.R., Tank, D.W.: Brain magnetic resonance imaging with contrast dependent on blood oxygenation. PNAS **87**, 9868–9872 (1990)
37. Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L.E., Hoke, M.: Increased auditory cortical representation in musicians. Nature **392**, 811–814 (1998)
38. Pantev, C., Roberts, L.E., Schulz, M., Engelien, A., Ross, B.: Timbre-specific enhancement of auditory cortical representations in musicians. NeuroReport **12**(1), 169–174 (2001)
39. Park, J.L., Fairweather, M.M., Donaldson, D.I.: Making the case for mobile cognition: EEG and sports performance. Neurosci. Biobehav. Rev. **52**, 117–130 (2015)
40. Petsche, H.: Approaches to verbal, visual and musical creativity by EEG coherence analysis. Int. J. Psychophysiol. **24**, 145–159 (1996)
41. Pylyshyn, Z.: Return of the mental image: are there really pictures in the brain? Trends Cogn. Sci. **7**(3), 113–118 (2003)
42. Révész, G.: Tonpsychologie. Voss, Leipzig (1913)
43. Rösler, F.: Statistische Verarbeitung von Biosignalen: Die Quantifizierung hirnelektrischer Signale. In: Baumann, U., et al. (eds.) Klinische Psychologie: Trends in Forschung und Praxis 3, pp. 112–156. Huber, Bern (1980)
44. Rouget, G.: Music and trance. A theory of the relations between music and possession. Chicago University Press, Chicago (1985)
45. Rumelhart, D.E., Norman, D.A.: Representation in memory. Stevens Handbook of Experimental Psychology 2, 2nd edn, pp. 511–587. Wiley, New York (1988)
46. Sagiv, N., Bentin, S.: Structural encoding of human and schematic faces: holistic and part-based processes. J. Cogn. Neurosci. **13**(7), 937–951 (2001)
47. Schneider, A.: Foundations of systematic musicology: a study in history and theory. In: Schneider, A. (ed.) Systematic and Comparative Musicology: Concepts, Methods, Findings, pp. 11–61. Peter Lang, Frankfurt am Main (2008)
48. Shepard, R.N., Metzler, J.: Mental rotation of three-dimensional objects. Science **171**, 701–703 (1971)
49. Siedentopf, C.M.: (Internet source) University of Innsbruck, Austria (2013). www.fMRI-easy.de
50. Sigalovsky, I.S., Melcher, J.R.: Effects of sound level on fMRI activation in human brainstem, thalamic and cortical centers. Hear. Res. **215**(1–2), 67–76 (2006)
51. Steinhauer, K., Alter, K., Friederici, A.D.: Brain potentials indicate immediate use of prosodic cues in natural speech processing. Nat. Neurosci. **2**(2), 191–196 (1999)

52. Stupacher, J., Hove, M.J., Novembre, G., Schütz-Bosbach, S., Keller, P.E.: Musical groove modulates motor cortex excitability: a TMS investigation. Brain Cogn. **82**, 127–136 (2013)
53. Talairach, J., Tournoux, P.: Co-Planar Stereotaxic Atlas of the Human Brain. 3-Dimensional Proportional System: An Approach to Cerebral Imaging. Thieme Medical Publishers, New York (1988)
54. Tayah, T.F., Abou-Khalil, B., Gilliam, F.G., Knowlton, R.C., Wushensky, C.A., Gallagher, M.J.: Musicogenic seizures can arise from multiple temporal lobe foci: intracranial EEG analyses of three patients. Epilepsia **47**, 1402–1406 (2006)
55. Tervaniemi, M., van Zuijen, T.L.: Methodologies of brain research in cognitive musicology. J. New Music Res. **28**(3), 200–208 (1999)
56. Thompson, T., Steffert, T., Ros, T., Leach, J., Gruzelier, J.: EEG applications for sport and performance. Methods **45**, 279–288 (2008)
57. Tiitinen, H., Virtanen, J., Ilmoniemi, R.J., Kamppuri, J., Ollikainen, M., Ruohonen, J., Näätänen, R.: Separation of contamination caused by coil clicks from responses elicited by transcranial magnetic stimulation. Clin. Neurophysiol. **110**, 982–985 (1999)
58. Wagemans, J., Vertraten, F.A.J., He, S.: Editorial—beyond the decade of the brain: towards a functional neuroanatomy of the mind. Acta Psychol. **107**, 1–7 (2001)
59. Warren, J.D., Uppenkamp, S., Patterson, R.D., Griffiths, T.D.: Separating pitch chroma and pitch height in the human brain. PNAS **100**(17), 10038–10042 (2003)
60. Zatorre, R.J., Belin, P.: Spectral and temporal processing in human auditory cortex. Cereb. Cortex **11**, 946–953 (2001)

Author Biography

Christiane Neuhaus studied Systematic Musicology at the University of Hamburg. She has strong research interests in neuromusicology and empirical musicology and was a postdoctoral fellow at the Max Planck Institute for Human Cognitive and Brain Sciences Leipzig for seven years. She wrote her Habilitation thesis about ‘Structural hearing’ and was recently invited as a guest professor at the Eberhard Karls University Tübingen. She presently works as a Privatdozentin at the University of Hamburg.

An Intelligent Music System to Perform Different "Shapes of Jazz—To Come"

Jonas Braasch, Selmer Bringsjord, Nikhil Deshpande, Pauline Oliveros and Doug Van Nort

Abstract In this chapter, we describe an intelligent music system approach that utilizes a joint bottom-up/top-down structure. The bottom-up structure is purely signal driven and calculates pitch, loudness, and information rate among other parameters using auditory models that simulate the functions of different parts of the brain. The top-down structure builds on a logic-based reasoning system and an ontology that was developed to reflect rules in jazz practice. Two instances of the agent have been developed to perform traditional and free jazz, and it is shown that the same general structure can be used to improvise different styles of jazz.

1 Introduction

Automated musical agents have a long tradition in Artificial Intelligence (AI) research. Starting first as composition tools [11, 17, 31, 18], modern computers are sufficiently fast to allow computational systems to improvise music with other performers in real time. Typically music composition/improvisation systems use a symbolic language, most commonly in form of the Musical Instrument Digital Interface (MIDI) format. Successful systems such as Lewis's Voyager system [20] and Pachet's *Continuator* [25] use MIDI data to interact with an individual per-

J. Braasch (✉) · N. Deshpande
Graduate Program in Architectural Acoustics,
Rensselaer Polytechnic Institute, 110 8th Street, Troy 12180, NY, USA
e-mail: braasj@rpi.edu

S. Bringsjord
Department of Cognitive Science, Rensselaer Polytechnic Institute,
110 8th Street, Troy 12180, NY, USA

P. Oliveros
Arts Department, Rensselaer Polytechnic Institute,
110 8th Street, Troy 12180, NY, USA

D.V. Nort
Computational Arts and Music, York University, Toronto, Canada

A. Schneider (ed.), *Studies in Musical Acoustics and Psychoacoustics*,
Current Research in Systematic Musicology 4,
DOI 10.1007/978-3-319-47292-8_12

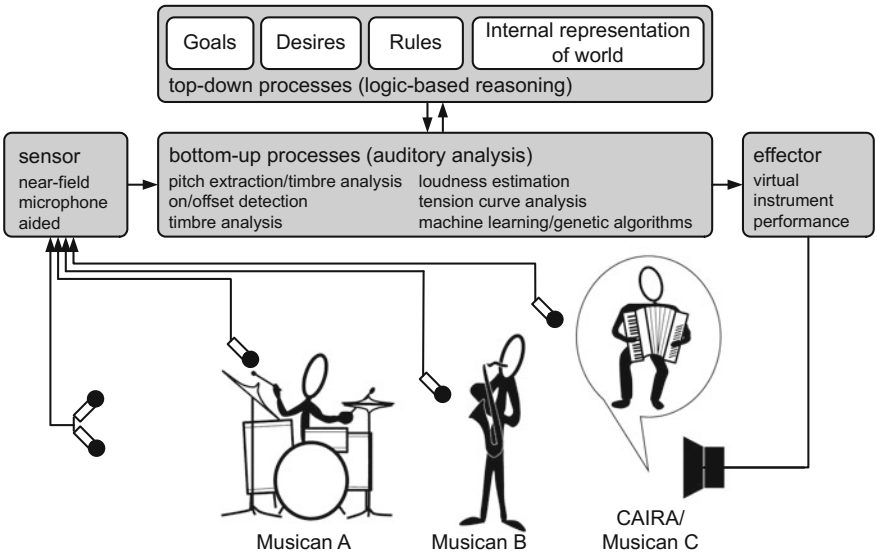


Fig. 1 Schematic of the creative artificially-intuitive and reasoning agent CAIRA

former whose sound is converted to MIDI using an audio-to-MIDI converter. The research described in this chapter stems from a larger project with the goal of developing a *Creative Artificially-Intuitive and Reasoning Agent* (CAIRA). Instead of using the simple audio-to-MIDI converter, the CAIRA uses standard techniques of Computational Auditory Scene Analysis (CASA) including pitch perception, tracking of rhythmical structures, and timbre and texture recognition (see Fig. 1). The CASA approach allows CAIRA to extract further parameters related to sonic textures and gestures in addition to traditional music parameters such as duration, pitch, and volume. This multi-level architecture enables CAIRA to process sound using bottom-up processes simulating intuitive listening and music performance skills as well as top-down processes in the form of logic-based reasoning. The low-level stages are characterized by a Hidden Markov Model (HMM) to recognize musical gestures and an evolutionary algorithm to create new material from memorized sound events. The evolutionary algorithm presents audio material processed from the input sound which the agent trains itself with during a given session, or from audio material that has been learned by the agent in a prior live session. The material is analyzed using the HMM machine listening tools and CASA modules, restructured through the evolutionary algorithms, and then presented in the context of what is being played live by the other musicians.

The logic-based reasoning system has been designed for CAIRA so it can "understand" basic concepts of music and use a hypothesis-driven approach to perform with other musicians (see top-down processes in Fig. 1). Including a logic-based reasoning system offers a significant number of benefits. The first goal is to see this multi-level approach lead to a more natural system response by trading off several

techniques; this makes the underlying processes less transparent to the human musicians without decreasing the overall responsiveness of the system. Secondly, the agent should be able to create new forms of music with the specific goal that the agent be able to develop its own concepts by expanding and breaking rules, and monitoring the outcome of these paradigm changes. Thirdly, we want to document the performance of the system—which is not easy to do—when the agent simulates intuitive listening in the context of Free Music. By adding a logic-based reasoning system, it is now possible to assess communication between the agent and human musicians by comparing the internal states of the agent and the human musicians.

This chapter focuses on the third goal for our logic-based reasoning stage. In particular, we describe a self-exploratory approach to test the performance of CAIRA within a trio ensemble. The approach, described in further detail below, is inspired by experimental ethnomusicology methods practiced by Arom [1] and others. A more detailed description of the lower- and higher-level CAIRA architecture and its ability to operate using the fundamental concepts of music ensemble interaction will precede this discussion.

1.1 Gestalt-Based Improvisation Model Based on Intuitive Listening

The artificially-intuitive listening and music performance processes of CAIRA are simulated using the *Freely Improvising, Learning and Transforming Evolutionary Recombination* (FILTER) system [28–30]. The FILTER system uses a Hidden Markov Model (HMM) for sonic gesture recognition, and it utilizes Genetic Algorithms (GA) for the creation of sonic material. In the first step, the system extracts spectral and temporal sound features on a continuous basis and tracks onsets and offsets from a filtered version of the signal. The analyzed cues are processed through a set of parallel Hidden Markov Model (HMM)-based gesture recognizers. The recognizer determines a vector of probabilities in relation to a dictionary of reference gestures. The vector analysis is used to determine parameters related to maximum likelihood and confidence, and the data is then used to set the crossover, fitness, mutation, and evolution rate of the genetic algorithm, which acts on the parameter output space [28].

1.2 Logic-Based Reasoning Driven World Model

One of the main goals of the CAIRA project was to understand how an artificially creative system can benefit from a joint bottom-up/top-down structure. CAIRA's knowledge-based system is described using first-order logic notation—for a detailed description of CAIRA's ontology see Braasch et al. [5]. For example, CAIRA knows that every musician has an associated time-varying dynamic level in

seven ascending values from *tacit* to *fortissimo*. The agent possesses some fundamental knowledge of music structure recognition based on jazz music practices. It knows what a solo is and understands that musicians take turns in playing solos while being accompanied by the remaining ensemble. The agent also has a set of beliefs. For example, it can be instructed to believe that every soloist should perform exactly one solo per piece.

One of the key analysis parameters for CAIRA is the estimation of the tension arc, which describes the currently perceived tension of an improvisation. In this context, the term 'arc' is derived from common practice of gradually increasing the tension until the climax of a performance is reached, and then gradually decreasing tension to end it. While tension often has the shape of an arc over time, it can also follow other trajectories. It is noteworthy that the focus here is not on tonal tension curves that are typically only a few bars long (i.e. demonstrating low tension whenever the tonal structure is resolved and the tonic appears). Instead, we are interested in longer structures, describing how a parameter relates to *Emotional Force* [22].

Using individual microphone signals, the agent tracks the running loudness of each distinct musical instrument using the Dynamic Loudness Model of Chalupper and Fastl [9]. The Dynamic Loudness Model is based on a fairly complex simulation of the auditory periphery that includes the simulation of auditory filters and masking effects. Additionally, the psychoacoustic parameters of roughness and sharpness are calculated according to Daniel and Weber [12] and Zwicker and Fastl [32]. In its current implementation, CAIRA estimates tension arcs for each musician from estimated psychophysical parameters. Based on these perceptual parameters and through its logic capabilities, the system recognizes different configurations for musical interplay. For example, it realizes that one of the musicians is performing an accompanied solo, by noticing that the performer is louder and has a denser texture than the remaining performers. The system can also notice that the tension arc is reaching a climax when all musicians perform denser ensemble textures. CAIRA takes action by either adapting its music performance to the analysis results or by presenting a dynamic visual score. CAIRA can, for example, suggest that a performer should end his or her solo because it is becoming too long, or it can encourage another musician to take more initiative. It can guide endings, and help an ensemble to fuse its sounds together.

Before we describe the mechanism to measure tension arcs, we briefly introduce the underlying basic concepts of jazz performance for two schools of jazz thought: traditional jazz, and free jazz.

2 Automated Music Improvisation Systems for Traditional Jazz

2.1 A Brief Overview on Traditional Jazz Practices

In this chapter, we use the term *traditional jazz* for jazz styles that precede the free jazz era—covering styles from swing to hardbop—but purposely exclude modal

jazz, which already contained numerous elements that later became characteristic features of free jazz. We will only cover the very basic fundamentals of jazz, but an extensive set of literature exists on this topic—for example Spitzer [27].

In traditional jazz, the freedom of an improviser is more constrained than one might think. Typically, each solo follows the chord progression of the song played by the rhythm section. The latter typically consists of drums, bass, and one or more chordal instruments—predominantly piano or guitar. For traditional reasons, one chord progression cycle is called a chorus.

The general repertoire of jazz tunes are called jazz standards. Most of these standards originated from Tin Pan Alley songs and pieces from Broadway musicals, in which jazz musicians performed for a living. After the theme is played the lead instruments take turns playing solos, and often players in the rhythm section take turns soloing as well. In traditional jazz, the performers are free to play over as many choruses as they want, but to end a solo before the end of the chord progression cycle is a taboo. The solo typically consists of a sequence of phrases that is chosen to match the chord progression and the intended dramaturgy. Since the two most common chord progressions in jazz are II-V and II-V-I (supertonic/dominant/tonic) combinations, professional jazz musicians train on phrases based on these progressions. Extensive literature exists with collections of standard jazz phrases.

Figure 2 shows the first eight bars of a notated saxophone solo over the 32-bar jazz standard, How High the Moon (Hamilton and Lewis 1940) to provide a practical example. Charlie Parker's Ornithology later used the same chord progression with a new bebop-style theme. Bars 3–6 consist of the typical II-V-I chord progression: G^{m7} (notes: G, B$^{\flat}$, D, F), C7 (C, E, G, B$^{\flat}$), F^{maj7} (F, A, C, E), and Bars 7 and 8 of another II-V progression Fm7 (F, A$^{\flat}$, C, E$^{\flat}$) and B$^{\flat 7}$ (B$^{\flat}$, D, F, A$^{\flat}$). Notice how in the example the saxophone initially follows the notes of the individual chords closely with additional scale-related notes—which is typical for swing. From Bar 6 on, the phrases change to bebop style with a faster eighth-note pattern. Also noteworthy is the second half of Bar 7, where the saxophone plays note material outside the chord related scale to create a dissonant effect. Whether this is appropriate depends on the agreed upon rules; In the swing era, this would

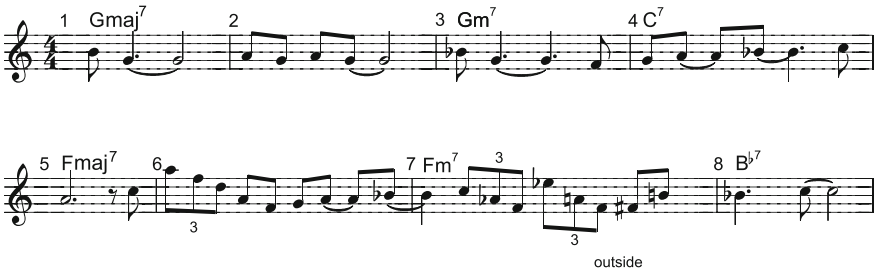


Fig. 2 Example transcription of a saxophone solo over the jazz standard How High the Moon (first 8 bars)—after Braasch [3]

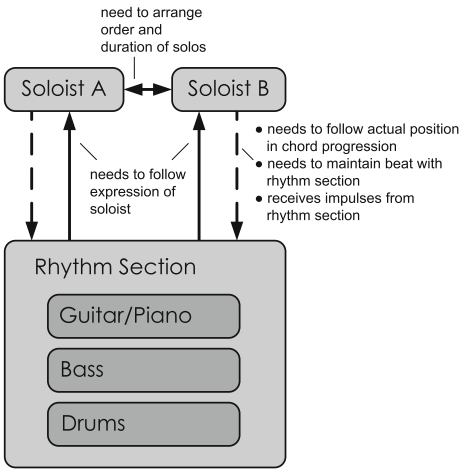


Fig. 3 Schematic communication scheme for a traditional jazz performance

have been played "incorrectly" but such techniques later became a characteristic style of players (such as Eric Dolphy) who could elegantly switch between so-called inside and outside play.

In order to play a correct solo following the rules of jazz theory, one could easily focus the attention to a very limited set of features to survive gracefully as shown in Fig. 3. Although, it should be noted that virtuoso jazz players are known to listen and respond to many details initiated by the other players. Basically, the soloist can process the rhythm section as a holistic entity, since all musicians follow the same chord progression. The tempo is quasi-stable, and the performance of the other soloist has to be observed only partially to make sure not to cut into someone else's solo. Once another soloist initiates a solo, he or she no longer needs to pay attention to the other soloists.

2.2 Rule-Based Machine Improvisation Algorithms

Numerous attempts have been made to design machine improvisation/composition algorithms to generate music material in the context of jazz and other styles [11, 17, 18, 31]. In most cases, these algorithms use a symbolic language to code various music parameters. The wide-spread MIDI (Musical Instrument Digital Interface) format, for example, codes the fundamental frequencies of sounds into numbers. Here, the note C_1 is the MIDI Number 24. Note numbers ascend in integers with the semitones. The temporal structure is also coded in numeral values related to a given rhythm and tempo structure.

By utilizing such a symbolic code, improvisation or composition can become a mathematical problem. Typically, the program selects phrases from a database according to their fit to a given chord progression (e.g., avoiding tones that are outside the musical scales for these chords, as previously discussed in context of

Fig. 2), and current position in the bar structure (e.g., the program would not play a phrase ending in the beginning of a chord structure). Under such a paradigm, the quality of the machine performance can be evaluated fairly easily by testing whether any rules were violated or not. Of course, such an approach will not necessarily lead to a meaningful performance, but the results are often in line with that of a professional musician. A system can even operate in real time as long as it has access to live music material on a symbolic level, for example, MIDI data from an electronic keyboard.

Lewis' Voyager system [20] and Pachet's Continuator [25] work using MIDI data to interact with an individual performer. The system transforms and enhances the performance of the human musician by generating new material from the received MIDI code, which can be derived from an acoustical sound source using an audio-to-MIDI converter; typically these systems fail if more than one musical instrument is included in the acoustic signal. In the case of the Continuator, learning algorithms are used based on a Hidden Markov Model help the system to copy the musical style of the human performer.

Commercial systems that can improvise jazz are also available. The program Band-in-a-Box™ is an intelligent automatic accompaniment program that simulates a rhythm section for solo music entertainers. The system also simulates jazz solos for various instruments for a given chord progression and popular music style. The system can either generate a MIDI score that can be auralized using a MIDI synthesizer, or create audio material by intelligently arranging prerecorded jazz phrases. The restricted framework of the jazz tradition makes this quite possible since the "listening" abilities of such a system can be limited to knowing the actual position within the form. Here the system needs to count along, making sure that it keeps pace with a quasi-steady beat.

3 Automated Music Improvisation Systems for Free Jazz

In contrast to traditional jazz, a formal set of rules does not exist in free jazz, although there has been a vivid tradition that has been carried on and expanded. Most of this tradition exists as tacit knowledge and is carried on in performance practice, orally and through musicological analyses. One example for tacit knowledge in free jazz is the taboo of performing traditional music material (see Jost [19]), unless it is a brief reference in the context of other adequate free music material. For the application of the informative feedback model to free jazz, it is also important to understand how the tradition progressed over time, deviating more and more from traditional jazz practice. A key moment for the development of free jazz was the introduction of modal jazz at the end of the 1950s, in which the chord progressions were replaced with fixed musical modes. In modal jazz the standard form of 12, 16 or 32 bars was initially kept, but this structure was given up in the favor of a free (variable) duration of form.

In the beginnings of free jazz, music material was fairly traditional and could be analyzed based on traditional music notation and thus easily captured using a symbolic music code like MIDI. As the field progressed musicians started to use extended techniques that shifted their performance more and more from the traditional sound production techniques of the orchestral instruments used in jazz. Albert Mangelsdorff's ability to perform multiphonics on the trombone is legendary, and so are the circular-breathed melodic streams of Evan Parker, who obtained the ability to perform arpeggio-style continuous phrases with a variable overtone structure containing both tonal and non-pitch-based elements. Peter Brötzmann's repertoire further expanded the techniques of non-pitched sounds. Among the younger generation of free jazz musicians are performers whose work focuses on complex musical textures outside the context of tonal music. Mazen Kerbaj (trumpet) and Christine Sehnaoui (saxophone) are among those who neglected the tonal heritage of their instruments in a unique way.

Initially, free jazz musicians took turns performing accompanied solos, but as time progressed it transformed into a genre where the boundaries between solos and accompaniment became blurred. While in traditional jazz a soloist has to listen to another soloist only to find a good slot for their own solo, instead performers began to pay attention all the time to other soloists. In addition, a soloist could no longer rely on the predetermined role of the rhythm section, which was now allowed to change keys, tempo and/or style. The higher cognitive load that was necessary to observe all other participants in a session led to smaller ensembles, often duos. Larger ensembles like the Willem Breuker Kollektief remained as the exception.

Figure 4 depicts a model of communication during a free jazz session. The diagram, shown here for a group of three musicians, appears to be much simpler because of the lack of rules. In contrast to the previous model for traditional jazz (Fig. 3), the distinction between rhythm section players and soloists is no longer made. While in traditional jazz the rhythm section can be represented as a holistic entity with homogeneous rhythm, tempo, and chord structure, now individual communication channels have to be built up between all musicians. Also, the feedback structure that each musician needs to enact to adequately respond to other players is fundamentally different from traditional jazz, where the communication

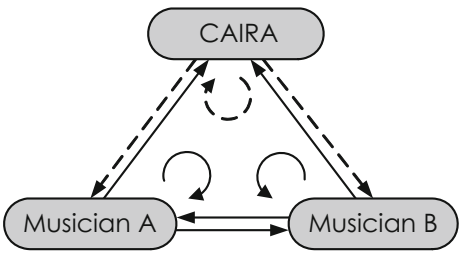


Fig. 4 Schematic of a communication scheme for a free jazz performance. The categorical distinction between soloists and rhythm section players no longer exists. Each musician has to establish individual communication channels to all other musicians, and also observe his or her own performance. The agents observation tasks are shown as *dashed arrows*

feedback loop (see Fig. 3) may simply cover a single communication stream from the ensemble (seen as a whole unit) to the soloist and back. In free music, a separate communication line has to be established between each possible pair of players, and consequently each performer has to divide his/her attention to observe all other players individually. Since the feedback from other musicians has to be detected with the ears, the multiple feedback-loop structure is not apparent in Fig. 1. However, the need to extract the information individually for each musician from a complex sound field is what makes free music improvisations a challenge. In addition, the performer always has to be prepared for unexpected changes, especially since tacit knowledge can be extended or modified within a session.

With regard to the music parameter space, for traditional jazz it is sufficient to receive the pitches of the notes played, to determine the current chord structure and melody lines, and to capture the onset and offset times of these notes to align the performance in time with the rhythm section. Commercial audio-to-MIDI converters can perform this task reliably enough if the general chord progression is known in advance. The analysis can even contain errors from information redundancy, as long as the algorithm can follow the given chord progression. In the context of an automated system that can improvise free music, machine listening demands are much higher if the system is mimicking human performance (see the Auditory Analysis box in Fig. 1). Here, it is no longer necessary to have a predetermined chord progression that serves as a general guideline. Even if there existed a system that could extract the individual notes from a complex chord cluster—which is difficult because of the complex overtone structure of the individual notes—it is not guaranteed that the musical parameter space in a session is based on traditional music notes.

To address this problem adequately, the intelligent system can be equipped with a complex model that simulates the auditory pathway. This type of model is able to extract features from acoustic signals in a similar way to the human brain (see Fig. 1). The early stages of the auditory pathway (auditory periphery, early auditory nuclei that perform spectral decomposition, pitch estimation, onset and offset detection) are thought to be purely signal driven, whereas the performance of the higher stages (e.g., timbre recognition, recognition of musical structure) are thought to be learned; these auditory features are categorized along learned patterns.

In the current CAIRA implementation, the features extracted in the Auditory Analysis stage are coded as symbolic information and passed on to the cognitive processing stage. From the symbolic information it receives, it can construct an internal representation of the world (in this case, the representation of the jazz performance). As outlined in the previous section, the art of mapping acoustic signals onto symbolic information is well defined through jazz theory for traditional jazz. Thus, if the system does not know and follows the given rules, it will be easily detected by other musicians and the audience.

In contrast, in free music, there is no longer a standardized symbolic representation of what is being played. Instead, to a greater degree, the music is defined by its overall sound. Consequently, the musicians will need to derive their own symbolic representation to classify what they have heard and experienced, and they

also need to define their own goals. For automated systems, the latter can be programmed using methods in second-order cybernetics (e.g., see Scott [26]). With regards to the symbolic music representation in humans, musicians typically draw from their own musical background, and significant differences can be found in musicians who primarily received classical music training compared to those who concentrated in jazz, or those that worked with sound textures rather than pitch and harmony. These differences extend to artists who learned in non-western music traditions. For example, if a traditionally trained musician hears a musical scale, they associate it with a scale that exists in their musical culture. This association works as long as the individual pitches of each note fall within a certain tolerance. Consequently, two people from two different cultural backgrounds could label the same scale differently, and thus operate in different symbolic worlds judging the same acoustic events. In free music, interactions between musicians of various cultural backgrounds are often anticipated, hoping that these types of collaborations will lead to new forms of music, and this precludes musicians falling into patterns. However, the communication will only work if the structures of different musical systems have enough overlap such that musicians can decipher a sufficient amount of features from other performing musicians into their own system. Furthermore, as performers, we have only indirect access to the listening ability of co-musicians through observing what they play, and in the case where something was not "perceived" correctly by others, we cannot measure their resulting response (musical action) along rules in free music, as these rules do not exist.

For cross-cultural music ensembles, examples exist where communication problems resulted from operating in different music systems. The late Yulius Golombeck once recalled when he was performing with the world music band Embryo, Charlie Mariano, and the Karnataka College of Percussion, there were certain complex Indian rhythms played by the Karnataka College of Percussion that the western trained musicians could not participate in because the rhythmical structure was too complicated to understand, despite the fact that all musicians had a tremendous experience with non-western music.[1]

While the complex communication structure in free music poses a real challenge for automated music systems, the lack of a standardized symbolic representation can be used to a system's advantage. Instead of mimicking the auditory system to extract musical features (Fig. 1), an alternative approach could be a robot-adequate design. The design could consider that as of today some parameters (e.g., complex chords) are impossible to extract in parallel for multiple musicians, especially in the presence of room reverberation. Instead, a music culture for machines can be developed that emphasizes the strengths of machines and circumvents their shortcomings. The latter is summarized in Table 1.

A directed focus on machine-adequate listening algorithms also encourages the design of machines to have their own identity, instead of focusing on making them indistinguishable from humans by passing the Turing test (e.g., compare Boden

[1]Braasch, personal communication, 1995.

Table 1 Listening strengths and weaknesses of machines compared to humans

Listening strengths	*Listening weakness*
• Absolute sense of time • Absolute sense of timbre • Absolute sense of pitch	• Difficulty to perceptually correct imperfections of other players • Difficulty to reconstruct missing information • Difficulty to extract information from multiple source and reverberant environments
Cognition strength	*Cognition weakness*
• Good at combinatorics • Absolute memory	• Has no understanding of aesthetics • Unable to develop new concepts • Cannot abstract ideas
Action strength	*Action weakness*
• Can play everything at any tempo • Redefines virtuosity	• Difficulty to adapt to other musicians • Difficulty to perform with musical expression

[2]). Man/machine communication can then be treated like a cross-cultural performance, where sufficient overlap between the various cultures is expected to allow meaningful communication. In such collaborations, the goal would not be to replace humans with machines, but to build systems that inspire human performers in a unique and creative way. A good example of machine inspired human music performance is the introduction of the drum machine, which encouraged a new generation of drummers around Dave Weckl in the 1980s to perform their instruments more accurately—almost in a machine-like style.

4 Implementation of CAIRA

In this section, we describe the different modules that were designed and implemented to operate the CAIRA agent. We first describe the bottom-up mechanisms, and then the top-down structures.

4.1 Bottom-Up Mechanisms

The bottom-up mechanisms are signal driven and include modules that simulate different functions of the auditory periphery including pitch detection [6], beat detection, loudness calculation and the calculation of tension curves [4, 8]. Further, the CAIRA system heavily uses machine learning algorithms—based on Hidden Markow Models (HMM) and Empirical Mode Decomposition to analyze sonic gestures based on different time scales. The machine learning algorithms, which are especially important for the Free Jazz Instantation of CAIRA, are subsumed in the FILTER structure and have been described thoroughly in peer-reviewed literature

[28–30]. We therefore focus on the description of the polyphonic pitch detection model and tension curve estimation in this chapter.

4.1.1 Polyphonic Pitch Perception Model

The polyphonic pitch model builds on a functional model of the auditory periphery and previous pitch perception models [10, 14, 21, 24]. In the first step, to simulate the behavior of the basilar membrane the signal is sent through a Gammatone filterbank with 128 bands to segregate sound into different auditory bands. Then, the signal frequency f_n in each band n is estimated using auto-correlation, measuring the delay τ_n between the main and the largest side peak:

$$f_n = \frac{1}{\tau_n}. \tag{1}$$

A novel aspect of the model is that both the frequency and pitch strength are measured in each frequency band, the latter calculated using the amplitude ratio a between the largest side peak and the main peak. Further, the deviation b between the estimated frequency f_n and the center of the frequency band $f_{c,n}$ is calculated. Next, all results are grouped into four categories:

1. $a > 0.90$, $b \leq 0.3$ octaves ('+' symbols)
2. $a > 0.90$, $b > 0.3$ octaves ('×' symbols)
3. $a \leq 0.90$, $b \leq 0.3$ octaves ('°' symbols)
4. $a \leq 0.90$, $b > 0.3$ octaves ('*' symbols)

The graphs on the left in Fig. 5 show the results of the pitch model for a 440-Hz sinusoid. The top graph shows the broadband autocorrelation function, the center graph the Fourier Transformation of the signal, and the bottom graph depicts the excitation of the auditory bands (solid black curve). For the curve, the energy in

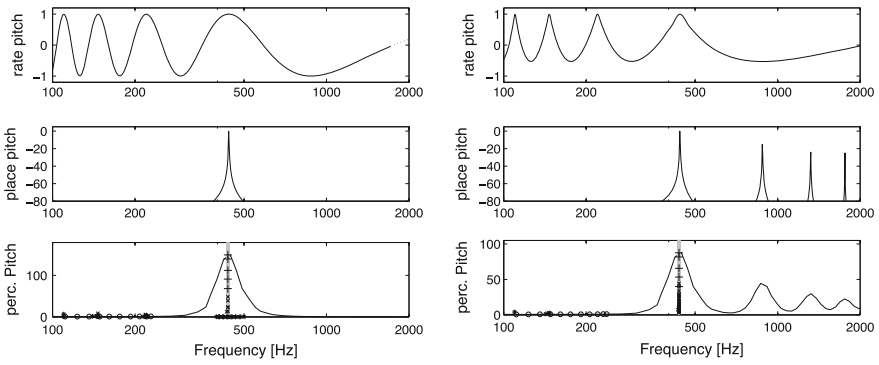


Fig. 5 Pitch estimation for a 440-Hz sinusoidal signal (*left graph*) and a 440-Hz tone complex (*right graph*)

each of the 128 bands was measured and plotted at the center frequency of the band. All values for the Group 1 are located at 440 Hz, the frequency of the sinusoid, as indicated by the gray curve. The height of the values represents the energy of the band in which the frequency was measured. The values for Group 2 also point to 440 Hz; they were measured in the adjacent side bands. All other values (Groups 3 and 4) were measured from the noise spectrum at low energies and do not represent the frequency of the sinusoid.

The right graphs of Fig. 5 depict the same context but this time for a tone complex with eight higher harmonics at integer multiples of the fundamental frequency: $f = n \cdot f_0$. The amplitude of the tone complex rolls off with $1/f$. Again, all values for Groups 1 and 2 ('+' and '×' symbols) point to the fundamental frequency of 440 Hz, even for those that belong to the higher harmonics.

Figure 6 shows the results for a $1/f$ tone complex at a lower fundamental frequency of 220 Hz (left graphs). Again, the results in all harmonics point to the fundamental, with the exception of two values in the octave region (440 Hz). It is not clear why in this case the octave is recognized; this will be further investigated. For higher harmonics, more than one overtone falls into the same auditory band. The overtones interfere with each other, and based on this interference the autocorrelation method identifies the common fundamental f_0. For the same reason, the algorithm is able to detect a missing fundamental. The right graphs show the results for the same tone complex, but this time the fundamental of 220 Hz was removed. Still, most values point to 220 Hz. Clearly, those values belong to Group 2 since there is no energy around 220 Hz and the values were computed for higher frequency bands.

Finally, chord complexes were analyzed using the model as depicted in Fig. 7. The left graph shows a triad of sinusoids with frequencies of 220, 262 and 330 Hz. The model correctly identifies all tones. The right graphs show a cluster of $1/f$ tone complexes with the following fundamental frequencies: 220, 262, 330 and 880 Hz. The model identifies all fundamental frequencies correctly, but also a number of octaves, for example at 516 Hz which is the octave of the 262-Hz tone.

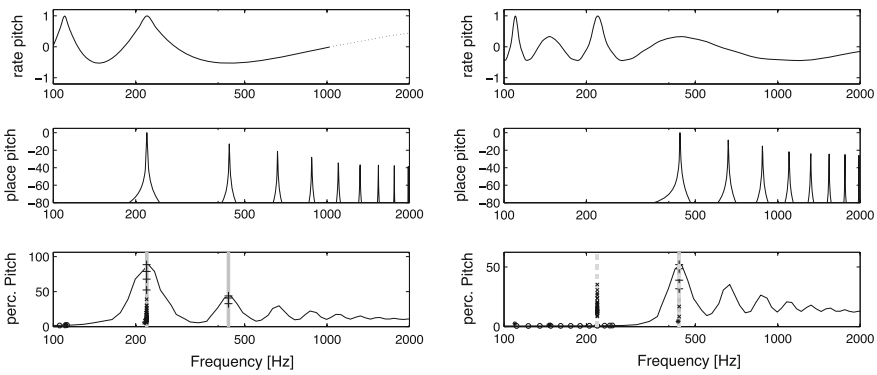


Fig. 6 Same as Fig. 5, but for 220-Hz signals

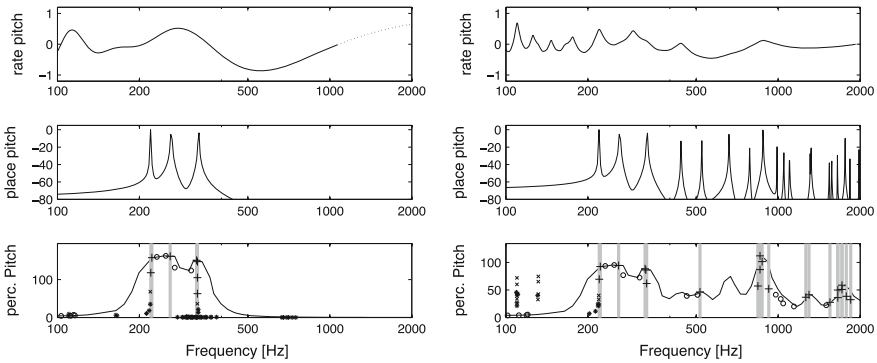


Fig. 7 Same as Figs. 5 and 6, but for polyphonic tone clusters. *Left* sinusoids with frequencies of 220, 262, and 330 Hz, *Right* tone complexes with frequencies of 220, 262, 330 and 880 Hz

4.1.2 Tension Arc Calculation

One of the key analysis parameters for CAIRA is the estimation of the tension arc, which describes the currently perceived tension of an improvisation. In this context, the term 'arc' is derived from the common practice of gradually increasing the tension until the climax of a performance section is reached, and then gradually decreasing tension to end it. Thus, tension often has the shape of an arc over time, but it can also have different time courses. It is noteworthy that we are not focusing here on tonal tension curves that are typically only a few bars long (i.e. demonstrating low tension whenever the tonal structure is resolved and the tonic appears). Instead, we are interested in longer structures, describing a parameter that is also related to Emotional Force [23].

Using individual microphone signals, the agent tracks the running loudness of each musical instrument using the Dynamic Loudness Model of [9]. The Dynamic Loudness Model is based on a fairly complex simulation of the auditory periphery including the simulation of auditory filters and masking effects. In addition, the psychoacoustic parameters of roughness and sharpness are calculated according to Daniel and Weber [12], and Zwicker and Fastl [32]. In the current implementation, CAIRA estimates tension arcs for each musician from simulated psychophysical parameters. Based on these perceptual parameters and its logic capabilities, the system recognizes different configurations for various patterns; e.g., it realizes that one of the musicians is performing an accompanied solo, by noticing that the performer is louder and has a denser texture than the remaining performers. The system can also notice that the tension arc is reaching a climax when all musicians perform denser ensemble textures. CAIRA takes action by either adapting its music performance to the analysis results or by presenting a dynamic visual score as described in more detail in the next section. CAIRA can, for example, suggest that a performer should end their solo because it is too long, or it can encourage another musician to take more initiative. It can guide endings and help an ensemble to fuse its sounds together.

In a previous study, we decided to calculate the tension arcs T from a combination of loudness L and roughness data R [5]:

$$T = L^4 + a \cdot R^3, \tag{2}$$

with an adjusting factor a. In a further study, we also suggested including information rate—e.g., as defined by Dubnov [15] and Dubnov et al. [16]—as an additional parameter for the tension arc calculation [7]. A real-time capable solution was developed to measure the rate and range of notes within each 2-s time interval. To achieve this, pitch is measured and converted to MIDI note numbers. Next, the number of notes n is counted within a 2-s interval, ignoring the repetition of identical notes. The standard deviation σ of the note sequence is then determined from the list of MIDI note numbers. Finally, the information rate I is determined from the product of *the number of notes* and *the standard deviation of MIDI note numbers*, or $I = n \cdot \sigma$. Practically, we measure values between 0 and 100. The tension curve is then calculated using the following equation:

$$T = \frac{1}{a+b}(a \cdot L + b \cdot ((1-q) \cdot R + q \cdot I)), \tag{3}$$

with the Information Rate I, Loudness L, and Roughness R. Note that all parameters, L, R, I, are normalized between 0 and 1 and the exponential relationships between the input parameters and T are also factored into these variables. The parameter q is the quality factor from the *YIN* pitch algorithm [14]. A value of one indicates a very tonal signal with a strong strength of pitch, while a value of zero indicates a noisy signal without defined pitch. The parameter is used to trade off roughness and information rate between tonal and noise-like signals. The parameters a and b are used to adjust the balance of loudness and the other input parameters for individual instruments. All tension curves are scaled integer values between zero and seven. Figure 8 shows an example of how a tension curve is estimated from the instruments' sound pressure signal.

4.2 Top-Down Mechanisms

A logic-based reasoning system is used to implement the top-down mechanism of CAIRA. The main purpose of the top-down mechanism of CAIRA was to provide the system with a rudimentary "understanding" of musical rules and concepts provided by a field-specific ontology. We hope that we will be able to expand the current architecture in the future such that CAIRA can use the externally injected knowledge to form its own concepts and ideas. The rule-based approach is also important for the system to be able to measure success by adhering to formal rules and "realizing" when these rules are broken. An important component of the CAIRA system is the interaction between the bottom-up and top-down

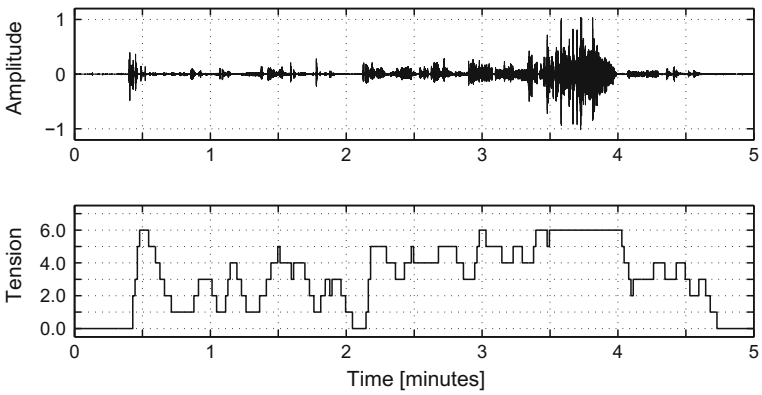


Fig. 8 Tension Arc calculation for a soprano saxophone sample. *Top* waveform of the saxophone, recorded with a closely positioned microphone. *Bottom* calculated tension arc curve—adapted from [4]

mechanisms. Of particular importance are the calculated tension arc curves, which are measured individually for each musician using closely positioned microphones, to "understand" the basics of musical interplay between the musicians. The rules of engagement for musical interplay are defined by the ontology that will be described in detail in the next section.

4.2.1 General Ontology Definitions

Our knowledge-based system is described using first-order logic notation.[2] We define that every musician has an associated dynamic level in seven ascending values from *tacit* to fortissimo, *ff*:

$$\begin{aligned}&[\forall x(Mx \rightarrow \forall t(Tt \rightarrow (d-l(x,t) = tacit \vee\ d-l(x,t)\\&\quad = pp \vee d-l(x,t) = p \vee d-l(x,t) = mp \vee d-l(x,t)\\&\quad = mf \vee d-l(x,t) = f \vee d-l(x,t) = ff)))]\\&\qquad \wedge (tacit < pp \wedge pp < p \wedge p < mp \wedge mp < mf \wedge mf < f \wedge f < ff).\end{aligned}$$

It is noteworthy here that the dynamic levels are calculated for every musician in discrete steps using the dynamic loudness model. The condition *tacit* is the case where the instrument does not produce a tone. Each moment in time is labeled through the variable S, with S_0, the start of the improvisation, and S_{END}, the end of

[2]Some of the equations have been simplified for better readability and for the reason of saving space.

the improvisation. If we have more than one musician, all existing musicians form an ensemble:[3]

$$\forall x \forall y((Mx \wedge My \wedge x \neq y) \rightarrow \exists z(Ez \wedge Ixz \wedge Iyz \wedge (\forall y(Ey \rightarrow y = z)).$$

In contrast, if we have less than two musicians an ensemble does not exist:

$$[(\neg \exists x Mx) \vee (\exists x(Mx \wedge \forall y(My \rightarrow y = x)))] \rightarrow \neg \exists z Ez.$$

4.2.2 Music Structure Recognition

Next, we define the current configuration of the ensemble. We divide the structure of the performance as a sequence of solos and ensemble parts, whereas each moment in time is characterized by a solo (of exactly one performer) or an ensemble part:

$$\begin{aligned} &\forall S(\neg \exists x Musician(x) \wedge PlaySolo(x, S)) \\ &\quad \Leftrightarrow (\forall x Musician(x) \wedge PlayEnsemble(x, S)). \end{aligned}$$

Obviously, the ensemble performance part cannot exist if we have only one performer. In this case, we can automatically conclude that the musician is performing a solo:

$$\forall x((Musician(x) \wedge \forall y(Musician(y) \Rightarrow x = y)) \Rightarrow PlaySolo(x)).$$

and the $PlayEnsemble(x, y, \ldots, S)$ mode does not exist:

$$\begin{aligned} &\forall x((Musician(x) \wedge \forall y(Musician(y) \Rightarrow x = y)) \\ &\quad \Rightarrow (PlaySolo(x) \wedge \neg PlayEnsemble(x))). \end{aligned}$$

Now, we have to decide how the agent recognizes the correct $PlaySolo(x, S)$, or the alternative $PlayEnsemble(x, y, \ldots, S)$, mode. First, our tension arc estimations for each musical instrument are based solely on their dynamic levels. This assumption needs to be refined at a later point, but high correlation values justify this initial approach. We define:

$$\forall x, S\ DynamicLevel(x, S) \Leftrightarrow TensionArc(x, S),$$

and leave it for later to refine the *TensionArc* calculation. Now we can define the solo performance mode as:

[3]Please note that we use the variable x and y for musicians, S for time, and z for an ensemble throughout this chapter.

$$\forall x, y, S\ (((Musician(x) \wedge TensionArc(x, y, S) \wedge \forall w, z \\ ((Musician(w) \wedge TensionArc(w, z, S)) \Rightarrow (y > z)))) \Rightarrow PlaySolo(x, S)).$$

Note that the solo performance mode relies on the fact that exactly one performer has to have a higher current tension arc value then all other performers. If at least two performers share the highest value, the agent recognizes the ensemble performance mode, which can be defined as:

$$\forall S\ ((\neg \exists x\ PlaySolo(x, S) \wedge \exists x\ (Musician(x) \\ \wedge \neg TensionArc(x, tacit, S))) \Rightarrow \forall x\ (Musician(x) \Rightarrow PlayEnsemble(x, S))).$$

The improvisation ends if the following condition is met:

$$\forall S\ (\forall x\ (Musician(x) \wedge TensionArc(x, tacit, S)) \Rightarrow EndOfMusic(S)).$$

We should reemphasize here that we calculate a running average dynamic level and not instantaneous values. The duration of the averaging window is crucial for the performance of the agent, but we observe similar challenges with human listeners. Take, for example, the case where an audience listens to an unknown classical composition. It always takes a certain time period until the first audience member decides when the piece is over and claps, and the audience often waits until the musicians bow. False alarms are often remembered as embarrassing incidents. Similarly, it is important that the tension arc has discrete values, otherwise, minimal tension arc differences between performers easily lead to the false detection of a solo. The integration of thresholds needs to be considered if the tension arc is calculated based on continuous values.

4.2.3 Agent Beliefs

Now we discuss the beliefs and goals of the agent. A simple example can be drawn from Jazz, where every soloist is expected to perform exactly one solo per piece:

$$\forall x, S\ (Musician(x) \wedge (NumberOfSolos(x, 0, S) \\ \vee NumberOfSolos(x, 1, S)) \Leftrightarrow DesiredState(S)).$$

In contrast, it is undesirable that a performer plays a second solo:

$$\forall x, S\ (Musician(x) \wedge PlaySolo(x, S) \wedge \neg NumberOfSolos(x, 0, S) \\ \Leftrightarrow TooManySolos(x, S)).$$

It is also an undesired state if the performer's solo gets too long:

$$\forall x, S\ ((Musician(x) \wedge PlaySolo(x, S) \\ \wedge SoloDuration(x, MaxSoloDuration, S)) \Rightarrow SoloTooLong(x, S)).$$

In this last aspect, there is much room for improvement in the agent's performance. Instead of simply assigning a threshold of what should be the maximum solo duration, the agent could observe if the performer is still producing interesting work or exhausted their creativity. An alternative method to determine if the solo becomes too long is:

$$\forall x, t, S\ [(Musician(x) \wedge PlaySolo(x, S) \wedge TensionArc(x, t, S) \\ \wedge (t < MinTensionArc) \wedge (t < MaxTensionArc)) \Rightarrow SoloTooLong(x, S)],$$

with S_{Solo} representing all moments in time of S during the performer's solo. Of course, in this case, the assumption that the tension arc simply relies on the dynamic level is very crude and the tension arc estimation should be refined, otherwise, musicians will be too restricted. Instead of simply observing the tension arc, the agent could also observe the variety of the performed solo material—e.g., via the information rate according to Dubnov et al. [16]—and the tension arc developments of the other musicians. The latter often declines, if the ensemble comes to the conclusion that the soloist should come to an end. The agent could also decide that the determination of whether a solo is too long is based on both the performance and a constant threshold. For example, John Coltrane was known to wear out the audience with long solos in Miles Davis' band, despite the excellent quality of his performance [13]. A good indicator that the solo of a performer will come to an end could be:

$$\forall x, t, S\ [(Musician(x) \wedge PlaySolo(x, S) \wedge TensionArc(x, t, S) \\ \wedge (t < AverageTensionArc)) \Rightarrow SoloMightEndSoon(x, S)],$$

which enables the agent to look ahead.

4.2.4 Action

Now we have to decide what action the agent should take if the ensemble reaches an undesired state (e.g., a solo is too long or a performer plays more than one solo within one piece). In a simple model, the agent can either

1. accept the undesired state
2. ask the other musicians ($x \neq y$) via a computer terminal if they find the solo to be too long. In this case, the agent can learn from their feedback and take further action (see below) based on the response
3. ask the performer to come to an end
4. encourage another musician to take the lead.

We can summarize this to the following proposition:

$$\begin{aligned}&\forall x, y, S[Solo\,TooLong(x, s)\\&\Rightarrow (AcceptUndesiredState(S)\\&\vee (AskMusiciansIfSolo\,TooLong(y, S) \wedge \neg(x = y))\\&\vee AskMusicianToStopSolo(x, s)\\&\vee (AskMusicianToplaySolo(y, S) \wedge \neg(x = y)\\&\wedge NumberOfSolos(y, 0.S)))].\end{aligned}$$

The agent can take similar measures if the performer plays a second solo and can also aid the musicians to end the improvisation if all performers have played a solo.

4.3 Implementation of a Free Jazz Agent

A Bayesian model is used to find an a posteriori estimation of the most likely ensemble state from the obtained tension curves. The ensemble states describe the instantaneous relationships between the musicians of an ensemble using methods in jazz ensemble practice. To keep the interaction sufficiently simple, we define six Ensemble States for a trio shown in the schematic in Fig. 4:

1. Solo A: Performer A performs a solo part
2. Solo B: Performer B performs a solo part
3. Solo C: CAIRA performs a solo part
4. Low-Tension Tutti: All ensemble members perform a tutti part with low tension
5. High-Tension Tutti: All ensemble members perform a tutti part with high tension
6. End: All musicians come to an end.

The Ensemble States are determined using a logic-based reasoning approach published in Braasch et al. [5], the practical rules that were derived in this study are given in Table 2. We cannot assume that each of the six states is performed equally long in time, but by using a Bayesian approach we can improve the Ensemble State estimation by recording how often each state occurs as a percentage over the whole training duration. To this purpose, the human performers use a foot pedal to update the Ensemble State. In addition, we can compare the states with instrumentally measured parameters. To see the general approach, let us focus on the analysis of the time-variant tension curves of Musicians A and B. We define seven discrete levels of Tension T. Curves will be computed for each participating musician and for CAIRA, so we have three tension curves: $(T_a(t), T_b(t), T_c(t))$. We can compute how often each tension level combination is observed for a given ensemble state:

Table 2 Ensemble state calculations based on logic-based reasoning

	Musician A	Musician B	CAIRA C
1 Solo A	$T_A + 1 > T_B$	$T_B - 1 < T_A$	$T_C - 1 < T_A$*
2 Solo B	$T_A - 1 < T_B$	$T_B + 1 > T_A$	$T_C - 1 < T_B$*
3 Solo C	$0 < T_A < 4$	$0 < T_B < 4$	Decision needed
4 Low Tension Tutti	$0 < T_A < 4$	$0 < T_B < 4$	Decision needed
5 High Tension Tutti	$T_B > 5$	$T_B > 5$	$T_B > 5$*
6 Ending**	$T_B = 0$	$T_B = 0$	$T_C = 0$*

The variables T_A, T_B, and T_C represent the tension curves of Musicians A, B, and CAIRA. The asterisks denote that CAIRA does not have to follow the suggestions by the other two musicians but can also respond by using a different tension curve level

$$P(E|T_{a,b}) = \frac{P(T_{a,b}|E)p(E)}{p(T_{a,b})}. \tag{4}$$

The parameter $T_{a,b}$ is the observed combined tension curve T for Musicians A and B. The Tension Curve T_c is not part of the analysis, since the intelligent agent CAIRA will observe the other two musicians to predict the current Ensemble State E. We have 49 discrete values for $T_{a,b}$, (7·7 Tension State combinations). The term $p(T_{a,b}|E)$ is the likelihood that the joint Tension Curve $T_{a,b}$ is observed for a given Ensemble State E. The term $p(E)$ is the probability that State E occurs independently of the tension curve status, and $p(T_{a,b})$ is the probability that the joint Tension Curve $T_{a,b}$ occurs independently of the ensemble state. Using the Equation given above we can compute the posterior estimate for each possible Ensemble State $E_1 - E_7$ for any Tension Curve pair $T_{a,b}$. An Ensemble State curve will be discussed further below (see also Fig. 9).

4.4 Implementation of a Traditional Jazz Agent

In this section, we describe a variation of CAIRA that is used to accompany a traditional jazz or popular music soloist instead of participating in a free improvisation. This version of CAIRA uses the aforementioned bottom-up/top-down algorithms to adjust an automated music accompany system to the live performance of a jazz soloist, for example, a trumpet player. For this purpose, the sound of the jazz soloist is captured with a microphone from a close distance, and musical features such as loudness, information rate, and musical tensions are extracted in real time. The extracted values are then used to control the probability that a certain accompany style is selected, and parameters like volume are adjusted. For example, if the soloist plays many musical notes within a short time frame (high information rate) it is much more likely that the rhythm section, performed by the CAIRA agent, will play in double time than is the case when the soloist performs a solo with only a few notes at a time.

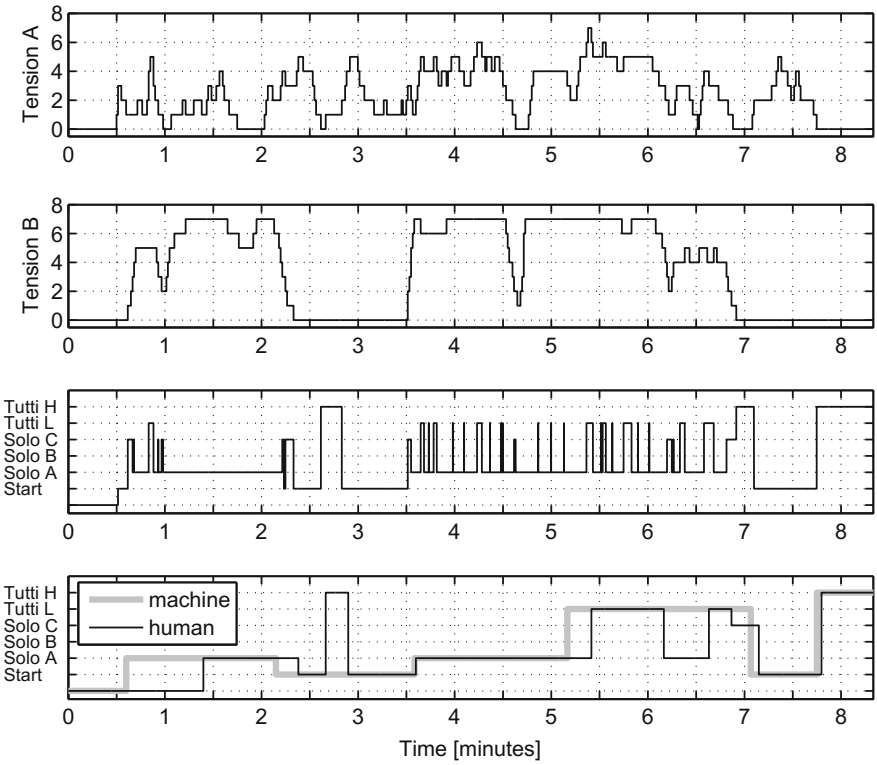


Fig. 9 Ensemble State Example for a trio session. *Top graph* tension curve for a saxophone (with variables a = 1.2 and b = 0.6); *2nd graph* tension curve for a Moog Synthesizer (a = 1.2, b = 0.4); *3rd graph* CAIRA's short term ensemble state estimations; *bottom graph* CAIRA's final (long-term) ensemble state estimations (*solid think black line*) versus human ensemble state estimations (*solid thick gray line*)—adapted from [4]

We believe that the traditional jazz version of CAIRA is particularly valuable from an educational point-of-view because it enables students to learn the jazz and popular music repertoire in a much more realistic setting than is the case for studying music with a pre-recorded backing tape because the agent offers interactive and dynamic system features.

Musical accompaniment systems have a long tradition in electronic organs used by one-man bands. Typically, the automated accompaniment produces a rhythm section (drums, bass, and a harmony instrument such as a piano) that performs in a given tempo (e.g., 120 beats-per-minute), style (e.g., Bossanova) and pre-defined chord progressions (often recorded live with the left hand of the organ player). The accompaniment system can then automatically generate a bassline and rhythmical harmonic chord structure from the performed chords and progressing chord structure. Similar systems, like Band-in-a-Box™, create a band that plays along from a manually entered chord sheet using software synthesizers for drums, bass and harmony instruments.

The problem with the current jazz/popular music accompany systems is that they are not "listening" to the performer, with the exception of systems that follow the tempo of the soloist.

Band-in-a-Box™, for example, will always perform a pre-rendered accompaniment that does not depend on the performance of the live soloist. In jazz, however, it is important that the players listen to each other and adjust their performance to the other players. For example, a good rhythm section will adjust its volume if the soloist plays with low intensity and uses sparse phrases. Often, some of the rhythm instruments rest and only part of the band accompanies the soloist. Or, the band can go into double time if the soloist plays rapidly (e.g., sequences of 16th notes). Double time is defined by playing at twice the tempo with each chord being performed twice as long in terms of musical measures such that the duration of the chord progression remains the same. In half-time, the tempo is half of the original tempo and the chord progression is half of the original metric value.

Impulses can also come from the rhythm section, for example, the rhythm section can decide to enter double time if the players believe an improvised solo could benefit from dynamic changes in structure. The adaptive performance of a rhythm section can become a real problem for a jazz student trying to practice unaccompanied. If the student is used to performing with a computerized rhythm section at home, then a live band changes this context dramatically. As a result, the jazz student is presented with a lack of experience for such situations as they may not be used to unexpected changes in the accompaniment. Likewise, it can become boring for even a highly experienced jazz player to perform with a virtual, static rhythm section that does not react to what is being played by the soloist.

Traditional jazz concretely defines the roles and basic groundwork for improvisation. In improvisation, solo instruments will introduce a composed theme and then introduce variations as elements of a solo. An improvised solo instrument will follow predetermined chord progressions that complete a phrase (such as the famous "twelve bar blues") accompanied by a rhythm section—traditionally drums, bass, and a chordal instrument such as piano or guitar. Figure 3 shows a diagram for how a traditional jazz group interacts. A standard solo will consist of an unspecified number of these complete chord progressions following the rhythm section. The more successful (and usually more famous) virtuoso jazz soloists are known to listen for and respond to details initiated by other members of their bands. In essence, the soloist sees the rhythm section as a holistic body following the chord progression, listening for details in melodic and rhythmic content from the section to incorporate into their solo. Tempo is usually held at a constant rate, and different lead instruments take cues from fellow musicians on when to introduce or fade out their own solo. In free jazz, individual solos soon blurred the line between soloists and accompaniment; in addition to distinguishing free jazz from its traditional roots, this process both increased the scope of a free jazz musician and added a layer of complexity to the improvisation process [27].

While there are rules in how to construct solos from traditional jazz theory, free improvisation gives the solo musician control over how such solos should evolve. However, as previously outlined, free improvisation carries its own history and

traditions. It seeks to eliminate almost all formal rule structures in composition. The introduction of modal jazz toward the late 1950s contributed greatly to free improvisation; chord progressions were replaced by musical modes, where instead of following pre-set changes composition instead pivoted from a musical tonic. Free improvisation at first kept the traditional structure of phrases, but soon progressed to free form in duration; this soon spilled over into other elements of performance. Popular techniques that stemmed from the experimentation of free improvisation included multiphonics, circular breathing, arpeggiated phrases, non-pitched tones, and complex textures discovered by approaching instruments in non-traditional methods [3].

Our traditional jazz agent listens to the soloist using a microphone—see Figs. 10 and 11. The system captures a number of acoustical and psychoacoustical parameters from the performed instrument including: (i) loudness, (ii) information rate (musical notes per time interval), and (iii) a tension curve based on loudness, roughness, and information rate. Alternatively, the system can compute these parameters directly from an electronic instrument (e.g., by analyzing MIDI data).

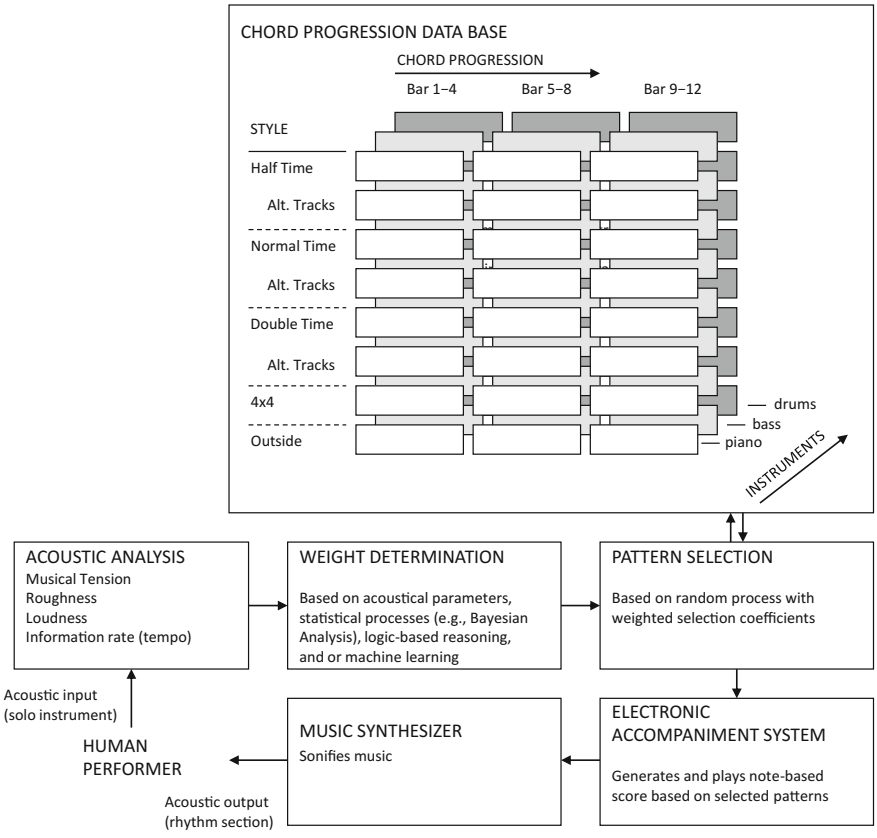


Fig. 10 System architecture for the CAIRA system for traditional jazz

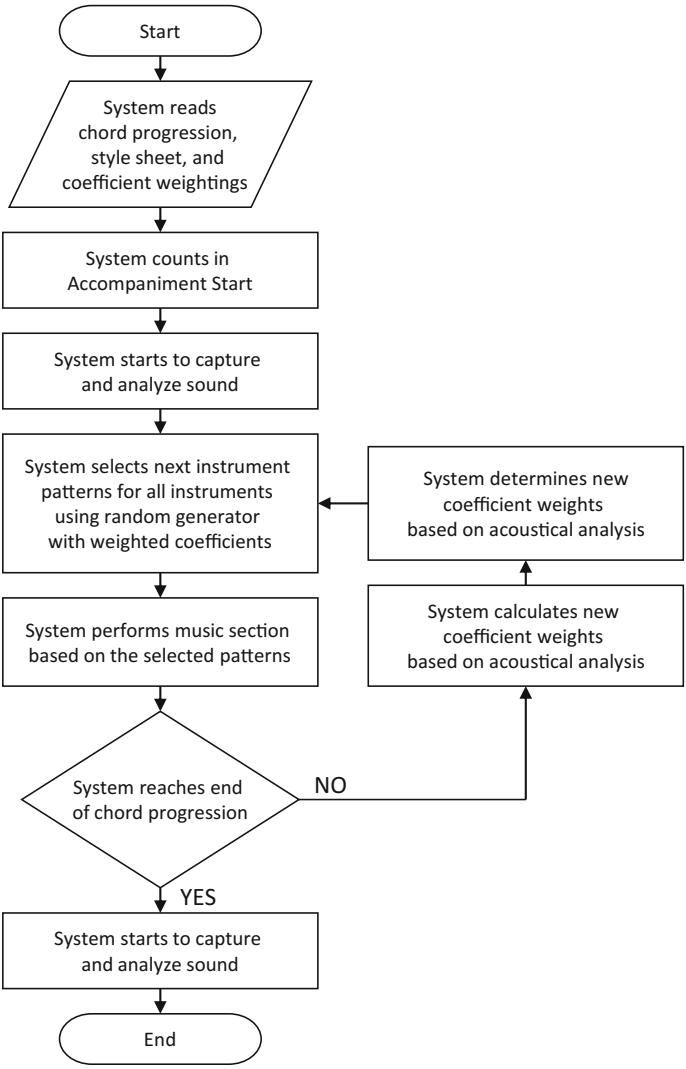


Fig. 11 System flow diagram for the traditional jazz-based CAIRA system

The accompaniment system then reacts to these measured parameters in real time making changes at strategic points in the chord progression (often at the end of four bars or the end of a phrase, or pre-specified chord structure). In particular the system will: (i) switch to double time if the soloists information rate and tension exceeds an upper threshold, (ii) perform at half time if the soloists information rate exceeds a lower threshold, (iii) return to normal time if the soloists information rate returns to in-between threshold rates, (iv) adapt the loudness of the rhythm section instruments to the loudness and tension curve of the performer, (v) play outside of the given chord structure if it detects the soloist performing outside this structure,

(iv) pause instruments if the tension curve or loudness is very low, or (vi) perform 4 × 4 between the solo instrument and a rhythm section instrument by analyzing the temporal structure of the tension curve (e.g., analyzing gaps or changing in 4-bar intervals). In a 4 × 4, the instruments take solo turns every four bars.

In addition, the rhythm section can give direction and take initiative based on a stochastic system using a random generator. For each event, a certain threshold of chance (likelihood) can be adjusted and if the internal drawn random number exceeds this threshold the rhythm section will take initiative in form of: (i) changing the style pattern, or taking a different pattern within the same style, (ii) stop instruments from changing to double time, half time, and normal time, (iii) lead into a new harmonic theme or other solos, (iv) play 4 × 4, and (v) play outside of expected structures. It should be noted that all changes can be subject to chance using a stochastic algorithm, for example by increasing the information rate to increase the likelihood for the rhythm section to change to double time, but there is no absolute threshold for these metrics.

5 Discussion and Conclusion

The purpose of this study was to develop a general framework for an intelligent music agent that can be adapted to different forms of music, in our case traditional jazz standards and free jazz. A dual bottom-up/top-down structure was chosen to simulate the creative processes needed to obtain a system that can perform live in an ensemble together with human musicians. The general architecture was identical for both types of jazz. Using the bottom-up structure an auditory scene analysis was performed, which included the estimation of pitch, loudness, information rate and beat among other parameters. A real-time tension curve was then calculated from these parameters to "understand" the intention of a soloist (traditional jazz agent) or to compare the inter-relationships between musicians (free jazz agent). A top-down structure, based on logic reasoning was used to control the agent according to specific rules of jazz.

One of the main goals for the dual bottom-up/top-down structure was to provide a mechanism where the system's response cannot be fully anticipated in advance, but at the same time to provide a framework where the human musicians who interact with CAIRA feel that the system is not responding in a random way, but "intelligently" responds to the performance of the live musicians. This can be achieved by tuning the parameter set to find the right balance between the genetic algorithms of the bottom-up stages (which can provide unexpected results) and the logic-based reasoning system, which can provide the feedback of being "understood".

One of the most critical aspects of the CAIRA project was to find solutions for the agent's ability to self-assess the success of its performance. Currently, the agent merely adheres to given rules or rejects to adhere to these rules, but it does not possess stages to assessing the musical structure according to aesthetical qualities. To achieve this is one of our long-term goals of the future.

Acknowledgments This material is based upon work supported by the National Science Foundation under Grant Nos. 1002851 and 1320059.

References

1. Arom, S.: The use of play-back techniques in the study of oral polyphonies. Ethnomusicology **20**, 483–519 (1967)
2. Boden, M.: The turing test and artistic creativity. Kybernetes **39**, 409 (2010)
3. Braasch, J.: A cybernetic model approach for free jazz improvisations. Kybernetes **40**(7/8), 972–982 (2011)
4. Braasch, J.: The μ-cos*m* project: an introspective platform to study intelligent agents in the context of music ensemble improvisation. In: Bader, R. (ed.) Sound–Perception–Performance, pp. 257–270. Springer, New York (2013)
5. Braasch, J., Bringsjord, S., Kuebler, C., Oliveros, P., Parks, A. Van Nort, D.: Caira—a creative artificially-intuitive and reasoning agent as conductor of telematic music improvisations. In: Proceedings of the 131st Convention of the Audio Engineering Society, Paper Number 8546 (2011)
6. Braasch, J., Oliveros, P. Van Nort, D.: Telehaptic interfaces for interpersonal communication within a music ensemble. In: 21st International Congress on Acoustics (ICA), Montreal, Canada (2013)
7. Braasch, J., Van Nort, D., Oliveros, P., Bringsjord, S., Govindarajulu, N. S., Kuebler, C. Parks, A.: A creative artificially-intuitive and reasoning agent in the context of live music improvisation. In: Music, Mind, and Invention Workshop: Creativity at the Intersection of Music and Computation, The College of New Jersey, Ewing, NJ. http://www.tcnj.edu/mmi/proceedings.html (2012)
8. Braasch, J., Van Nort, D., Oliveros, P., Bringsjord, S., Sundar Govindarajulu, N., Kuebler, C. Parks, A.: Music, Mind, and Invention Workshop: Creativity at the Intersection of Music and Computation, The College of New Jersey (2012)
9. Chalupper, J., Fastl, H.: Dynamic loudness model (DLM) for normal and hearing-impaired listeners. Acta Acustica United Acustica **88**, 378–386 (2002)
10. Cheveigné, A.: Pitch perception models. In: C. J. Plack A. J. Oxenham (eds.) The Psychophysics of Pitch. Springer, Berlin, Heidelberg, New York, pp. 169–233 (2005)
11. Cope, D.: An expert systemfor computer-assisted composition. Comp. Music J. **11**(4), 30–46 (1987)
12. Daniel, P., Weber, R.: Psychoacoustical roughness: implementation of an optimized model. Acustica **83**, 113–123 (1997)
13. Davis, M., Troupe, Q.: Miles, the Autobiography. Simon & Schuster, New York (1990)
14. de Cheveigné, A., Kawahara, H.: Yin, a fundamental frequency estimator for speech and music. J. Acoust. Soc. Am. **111**, 1917–1930 (2002)
15. Dubnov, S.: Non-gaussian source-filter and independent components generalizations of spectral flatness measure. In: Proceedings of the International Conference on Independent Components Analysis (ICA2003), pp. 143–148. Porto, Portugal (2003)
16. Dubnov, S., McAdams, S., Reynolds, R.: Structural and affective aspects of music from statistical audio signal analysis. J. Am. Soc. Inform. Sci. Technol. **57**(11), 1526–1536 (2006)
17. Friberg, A.: Generative rules for music performance: a formal description of a rule system. Comput. Music J. **15**(2), 56–71 (1991)
18. Jacob, B.: Algorithmic composition as amodel of creativity. Organ. Sound **1**(3) (1996)
19. Jost, E.: Free Jazz. DaCapo, New York (1981)
20. Lewis, G.: Too many notes: computers, complexity and culture in voyager. Leonardo Music J. **10**, 33–39 (2000)
21. Licklider, J.: A duplex theory of pitch perception. Cell. Mol. Life Sci. **7**(4), 128–134 (1951)

22. McAdams, S., Smith, B., Vieillard, S., Bigand, E. Reynolds, R.: Real-time perception of a contemporary musical work in a live concert setting. In: Stevens, C., Burnham, D., McPherson, G., Schubert, E., Renwick, J. (eds.) Proceedings of the 7th International Conference on Music Perception and Cognition, Sydney, Australia (2002a)
23. McAdams, S., Smith, B., Vieillard, S., Bigand, E. Reynolds, R.: Real-time perception of a contemporary musical work in a live concert setting. In: 7th International Conference on Music Perception and Cognition, Sydney, Australia, vol. 17, p. 21 (2002b)
24. Meddis, R., O'Mard, L.: A unitary model of pitch perception. J. Acoust. Soc. Am. **102**, 1811 (1997)
25. Pachet, F.: Beyond the cybernetic jam fantasy: the continuator. IEEE Comput. Graph. Appl. **24**, 31–35 (2004)
26. Scott, B.: Second-order cybernetics: an historical introduction. Kybernetes **33**(9/10), 1365–1378 (2004)
27. Spitzer, P.: Jazz Theory Handbook. Mel Bay Publications, Pacific, MO (2001)
28. Van Nort, D., Braasch, J., Oliveros, P.: A system for musical improvisation combining sonic gesture recognition and genetic algorithms. In: Proceedings of the 6th Sound and Music Computing Conference, pp. 131–136. Porto, Portugal (2009)
29. Van Nort, D., Braasch, J. Oliveros, P.: Mapping to musical actions in the filter system. In: The 12nd International Conference on New Interfaces for Musical Expression (NIME), Ann Arbor, Michigan (2012)
30. Van Nort, D., Oliveros, P. Braasch, J.: Developing systems for improvisation based on listening. In: Proceedings of the 2010 International Computer Music Conference (ICMC 2010), New York, NY (2010)
31. Widmer, G.: Qualitative perception modeling and intelligent musical learning. Comput. Music J. **16**(2), 51–68 (1992)
32. Zwicker, E., Fastl, H.: Psychoacoustics: Facts and Models, 2nd edn. Springer, Berlin (1999)

Author Biographies

Jonas Braasch is a psychoacoustician, aural architect, and experimental musician. His research work focuses on functional models of the auditory system, large-scale immersive and interactive virtual reality systems, and intelligent music systems. Jonas Braasch received a Master's Degree in Physics from the Technical University of Dortmund in 1998, and two doctoral degrees from the University of Bochum in Electrical Engineering and Information Technology in 2001 and Musicology in 2004. Afterward, he worked as Assistant Professor in McGill University's Sound Recording Program before joining Rensselaer Polytechnic Institute in 2006, where he is now Associate Professor in the School of Architecture and Director of the Center for Cognition, Communication, and Culture.

Selmer Bringsjord is a full professor in the Department of Cognitive Science at Rensselaer Polytechnic Institute. He teaches artificial Intelligence (AI), formal logic, human and machine reasoning, and philosophy of AI. Bringsjord's education includes a B.A. in Philosophy from the University of Pennsylvania, and a Ph.D. in Philosophy from Brown University. He conducts research in AI as the director of the Rensselaer AI & Reasoning Laboratory (RAIR). He specializes in the logico-mathematical and philosophical foundations of AI and cognitive science, and in collaboratively building AI systems on the basis of computational logic.

Nihil Deshpande is a doctoral student in architectural acoustics at Rensselaer Polytechnic Institute. His focus is on audio digital signal processing, specifically on computational auditory scene analysis and computer models of binaural hearing. Nikhil specializes in understanding,

developing, and extending digital audio effects, particularly reverberation algorithms. His master's research included a polyphonic pitch decomposition model and a dynamic real-time jazz accompaniment system. He has also been a programmer on Pauline Oliveros's Expanded Instrument System, for which his work was featured in the main gallery of the Whitney Museum of American Art as part of their Biennial in 2014. He received his B.S. in Electrical Engineering with a minor in Electronic Arts in 2013, and his M.S. in Architectural Acoustics in 2014, both from RPI.

Pauline Oliveros is a senior figure in contemporary American music. Her career spans fifty years of boundary dissolving music making. In the '50s she was part of a circle of iconoclastic composers, artists, poets gathered together in San Francisco. Recently awarded the John Cage award for 2012 from the Foundation of Contemporary Arts, Oliveros is Distinguished Research Professor of Music at Rensselaer Polytechnic Institute, Troy, NY, and Darius Milhaud Artist-in-Residence at Mills College. Since the 1960's she has influenced American music profoundly through her work with improvisation, meditation, electronic music, myth, and ritual. Pauline Oliveros is the founder of the Deep Listening Institute, now the Center For Deep Listening at Rensselaer.

Doug Van Nort is a computer music researcher and electronic music composer/improviser whose work is concerned with distributed agency, computational creativity, electroacoustic improvisation and sensorial immersion in technologically-mediated environments. He is Canada Research Chair in Digital Performance and an Assistant Professor in Computational Arts and Music at York University in Toronto.

Explorations in Keyboard Temperaments. Some Empirical Observations

Albrecht Schneider and Andreas Beurmann

Abstract In this article, the topic of tuning and temperament is addressed mainly from an empirical point of view. After furnishing some historical background on tone systems, scales, and tunings (in a review from Greek antiquity to the 18th century), twelve major and twelve minor chords played in two well-known keyboard tunings and temperaments (Werckmeister III, Vallotti) are investigated in regard to acoustical parameters on the basis of sound recordings we made with a Kirckman harpsichord from 1766. Our analysis of major and minor chords employs signal processing methodology, in particular autocorrelation and crosscorrelation from which the harmonics-to-noise ratio (HNR) is computed in the time domain as a measure of the periodicity in a signal. HNR readings vary for different chords relative to the justness of interval ratios and the different degrees to which partial frequencies converge in signals representing several complex harmonic tones such as contained in musical chords. The HNR thus can be taken as an indicator for the relative quality of a particular tuning. In addition, data from two experiments are reported in which listeners judged perceptual qualities as well as the goodness of intonation for various tunings implemented on digital synthesizers or realized by means of a computer. Our study intends to provide empirical data that can help to substantiate discussions of musical tunings and temperaments.

1 Introduction

Tuning of keyboard instruments to certain scale types and temperaments has been an issue for organologists and musicologists since long. In the past decades, surviving instruments have been investigated with the aim to possibly determine and reconstruct their original tuning. This approach proved effective in particular for

A. Schneider (✉)
Institut für Systematische Musikwissenschaft, Neue Rabenstr. 13, 20354 Hamburg, Germany
e-mail: aschneid@uni-hamburg.de

A. Beurmann
Institute of Systematic Musicology, University of Hamburg, Hamburg, Germany

A. Schneider (ed.), *Studies in Musical Acoustics and Psychoacoustics*,
Current Research in Systematic Musicology 4,
DOI 10.1007/978-3-319-47292-8_13

organs of the Baroque era where, for example, the original meantone tunings could be determined from measuring pipe lengths and diameters (see [1]). In regard to harpsichords and clavichords, a number of instruments (primarily from Italy) were found that had more than twelve keys to the octave originally, and were reduced to the conventional format later (see [2]). In addition to surviving instruments and related sources (such as drawings of keyboards), there is of course a huge body of theoretical works in which tone systems, scale types and modes as well as aspects of tuning and intonation are treated from Greek antiquity through the Middle Ages, the Renaissance, and then through modern times up to the present (see, e.g., [3–14]). In addition to theoretical writings, there are of course many musical works which reflect certain ideas about tone systems and modal structures, and which offer also clues in regard to tunings and intonation practice. It is from the analysis of musical works that conclusions may be drawn as to intended tunings (in particular on keyboards; see, e.g., [3, 8, 12, 15–18]).

With an increased interest in organology as well as in historical performance practice of Renaissance and Baroque music in the 20th century, tuning and temperaments gained also practical importance. One outcome of this process was that on a significant number of extant historical organs in Europe their original meantone tuning was reinstalled or that one of the well-tempered tuning systems (such as proposed by Werckmeister [19], Kirnberger [20]) were implemented in a tentative reconstruction of tunings in use before ET12 became the standard (in close connection with the development of the modern piano). Another factor is that harpsichords are now tuned to a rather low pitch (with A_4 often in the range from 385 to 408 Hz) while historical organs are re-tuned to their original pitch, which for many instruments was set by the *Chorton* (church tone, ton de chapelle, etc.) that was in use in a certain region. For instance, in Northern Germany the Chorton in use in the 17th century was about one semitone to two semitones higher than $A_4 = 440$ Hz (in other regions of Europe, it was almost equal to, or lower than the modern A_4 standard pitch).

With the revival of historical tunings and temperaments, also discussions concerning the merits and shortcomings of particular tunings and temperaments have been revitalized. Readers familiar with historical sources from the 16th, 17th, and 18th century, respectively, will recall that many proposals for temperaments and tunings aimed at providing a tonal basis for harmonic modulation through many keys while pleasing the musical ear (for background information, see e.g. [3, 8, 12, 18]).

2 Just Intervals: Acoustic and Perceptual Aspects

Humans (and apparently also other mammals) perceive two sine tones whose frequency ratio is 2:1 as similar in certain respects. The interval these tones form in music is labelled octave since it comprises, in many musical cultures, a scale of eight tones or notes (in this article, the term *tone* denotes a physical phenomenon

while *note* relates to musical notation. Of course, a musical note, say A_3, when played or sung as sound becomes a physical phenomenon as well as a psycho-physiological phenomenon in regard to sensation and perception). Perfect octaves have a distinct quality (which is restricted to the sense of hearing and to auditory perception) since their constituents match in a specific temporal and spectral pattern (see [21]). Likewise, just intervals such as the fifths (3/2) express a clear temporal and spectral structure. For two harmonic complexes each comprising a fundamental frequency, f_1, as well as harmonics f_2, f_3,…, f_n with a spectral envelope where amplitudes decay in a regular pattern like $A_n = 1/n$, the resulting signal is strictly periodic with a period $T = 1/f_0$ as is obvious from Fig. 1. The two fundamental frequencies here are $f_{1a} = 200$ Hz and $f_{1b} = 300$ Hz, and the frequency f_0 (plotted in red) with which the complex waveshape repeats is 100 Hz, that is, $T = 10$ ms.

Strict periodicity in the time domain corresponds to strict spectral harmonicity in the frequency domain. According to theorems developed by Wiener [22] and by Khintchine [23], the power spectrum W(ω) of a stationary time function f(t) equals the Fourier transform of its autocorrelation function φ(τ). Hence, for a periodic signal the autocorrelation function must also be periodic. The theorems of Wiener and Khintchine have been fundamental to the theory of vibration as they relate the concepts of time function and spectrum in regard to periodicity and harmonicity (see [24, 25]). In signals such as musical sound, one can easily see that a periodic vibration pattern observed, for example, from a thin string of a harpsichord, produces a highly harmonic spectrum where $f_n = nf_1$, $n = 1, 2, 3,\ldots, k$, that is, partial

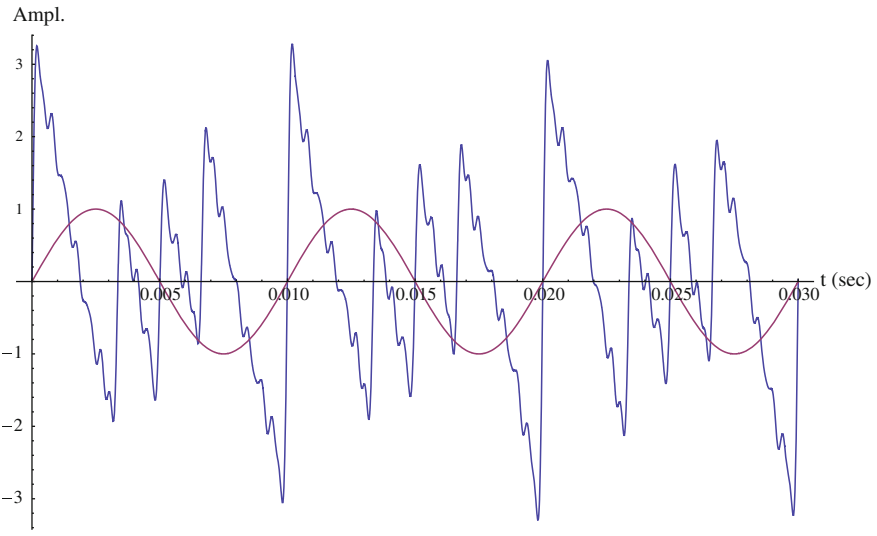
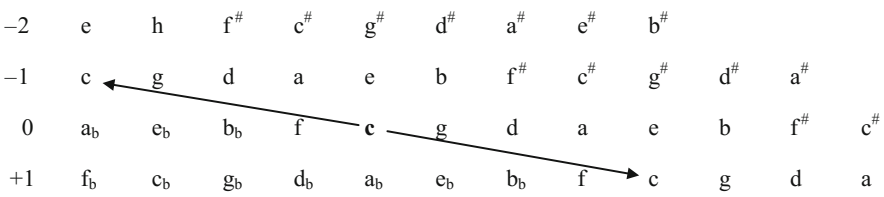


Fig. 1 Perfect fifth, two harmonic complexes, each comprising 10 harmonics, amplitudes $A_n = 1/n$, fundamental frequencies at 200 and 300 Hz, three periods shown. F_0 (repetition frequency of the complex waveshape) marked in *red*

frequencies are in integer frequency ratios (or very nearly so since inharmonicity from stiffness in thin brass strings of a harpsichord is almost negligible, see [26, 27]). The relation between temporal periodicity and spectral harmonicity defined by the theorems of Wiener and Khintchine is of central importance also for auditory perception of pitch and timbre (see [21]). Perceptual salience observed for just intervals can be explained by their high degree of periodicity and harmonicity, respectively, which for pairs of complex harmonic sounds played simultaneously implies a high degree of coinciding partials and, consequently, a low degree of roughness and beats (see [28, 29]). In addition, combination tones as well as perception of the 'missing fundamental' (see [30, 31]) come into play. For the perfect fifth shown in Fig. 1, when set to sound and played with sufficient level, a pitch component at 100 Hz will be clearly audible resulting from both f_0 (which is the repetition frequency of the period that furnishes a 'virtual' fundamental at 100 Hz to f_{1a} = 200 Hz and f_{1b} = 300 Hz) and the difference tone $f_{1b} - f_{1a}$. It is because of these facts which are open to empirical research that humans around the globe opted for musical intervals like the octave, the fifth (and its complementary interval, the fourth) as the most elementary (and most stable) building blocks for tone systems and scales. A good case in point is the scale for anhemitonic pentatony comprising five tones derived from a progression in fifths like c–g–d–a–e → c–d–e–g–a. Anhemitonic pentatony is found in very many music cultures (and may be viewed as a 'near universal' in music).

However, a fundamental problem behind the construction of tone systems and scales is that a finite sequence of just fifths $(3/2)^n$ will not form a cycle (but will take the shape of a spiral instead, see [32]). Taking a series of 12 fifths (e.g., from b_b–f–c–g… to $a^{\#}$), their compound size (which adds up to 8424 cents) overshoots that of seven octaves (8400 cents) by nearly 24 cents. The difference is known as the Pythagorean comma. The mathematical problem stated as $3^n \neq 2^m$ says that powers of one prime number do not equal powers of another prime number. For this reason, also three just major thirds of the ratio 5/4, when added to one interval (e.g., c–e–$g^{\#}$–$b^{\#}$), do not match a full octave 2/1 since their ratio of 125/64 falls short of that interval by about 41 cents (the gap corresponding to an interval of the ratio 128/125 = 1.024; this interval is called, close to classical Greek theory, a diësis). Again, the problem is that a tuning process which involves a series of three just major thirds would not yield an octave since $5^n \neq 2^m$. In effect, a *k*-dimensional tone net or tone lattice results from tone systems based on intervals each of which includes a prime number like 3/2 in the perfect fifth and 5/4 in the just major third. If the tone net represents the perfect fifth on the horizontal axis and the just major third on the vertical as the two fundamental intervals, the tone lattice or tone-net is a plane (as was explored first by Euler, and later by Arthur von Oettingen, Adriaan Fokker, and Martin Vogel). In case the 'natural' seventh 7/4 and thus the prime number 7 is included, the tone net is three-dimensional (see [14]).

A small segment (chosen to avoid double sharps and double flats) from the two-dimensional tone-lattice incorporating perfect fifths and just major thirds would be this:



Obviously, there are tones which have the same designation but appear in different rows of the plane. The number −1 indicates that tones in this row are flat by one so-called syntonic comma against the tone of the same name in the basic row (0). The syntonic or 'third' comma (ascribed to the Hellenistic music theorist Didymos) is the difference between two whole tones 9/8 and a just major third like (9/8) * (9/8) * (4/5) = 81/80 = 21.5 cents. For example, the just major third e (5/4) over c (1/1) is one syntonic comma flat against the Pythagorean *ditonos* e (81/64) derived from a progression in perfect fifths c–g–d–a–e. In the scheme of the tone-net sketched above, the tone c in the −1-row is one syntonic comma flat against the c in the 0-row (taken as a centre and marked in bold) while the c in the +1-row is one comma sharp (the tones **c**, c^{-1} and c^{+1} are marked by arrows). To play a chord of c-major in just intonation would require the tones c and g from the 0-row and the tone e from the −1 row (designated e^{-1} or $\underline{e}$). Likewise, a c-minor chord played in just intonation would need the tones c and g from the 0-row and the e_b (designated e_b^{+1} or $\bar{e}_b$) from the +1-row.

Just intonation based on intervals of the perfect fifth and fourth as well as on just major and minor thirds permits to render major and minor chords with a maximum of auditory consonance and thus a minimum of roughness and beats. What is more important, though, is that chord progressions in tonal harmony can be rendered so that truly chromatic and enharmonic textures become audible (and can be appreciated by listeners as complex pitch and interval structures). The cost for this achievement is that, first of all, far more than 12 tones and pitches per octave are required in particular for extended harmonic modulations. Furthermore, a problem can arise if extended modulations lead to chord structures that require tones far away from the centre of the tone-net. In such instances, the pitch level can shift by several commas (see [3, 33, 34]). Of course, in practice one may define a limit from where a 'reset' towards the centre takes place (see [14]). One may also limit the number of just intonation pitches which are implemented, in a fixed tuning on a pipe organ or electronic keyboard instrument, by making a selection of the musically most important tones and intervals. This was the approach chosen by the Norwegian composer and music theorist, Eivind Groven, for a pipe organ which had 36 tones and pitches to the octave, and for an electronic keyboard comprising 43 pitches to the octave (see [35, 36]).

3 Just Intonation and Temperaments: A Brief Historical Review

The theory of just intonation, which has origins in Greek and Hellenistic antiquity (see [14]) stems from both mathematical considerations and empirical observation. In regard to the former, divisions of integer ratios into smaller units were of relevance. Well known are divisions of the tetrachord where the frame of a perfect fourth 4/3 can be divided into three intervals in various ways (yielding either a diatonic, or a chromatic, or an enharmonic tone and interval structure). The so-called Diatonon ascribed to the theorist Didymos (1st century) and the Diatonon syntonon of Claudius Ptolemaios (2nd century) both divide the fourth into a major and minor whole tone, leaving a diatonic semitone as a rational (superparticular) interval: $4/3 = 9/8 \times 10/9 \times 16/15$ (Didymos) and $4/3 = 10/9 \times 9/8 \times 16/15$ (Ptolemy). This division implies the just major third 5/4 since 9/8 * 10/9 = 5/4. Apparently, the just major third was known to Greek theorists since Archytas of Tarent (4th century B.C.E.). As Ptolemy (ed. Düring 1934, 30f. [37]) asserts, Archytas calculated the diatonon, the chroma, and the enharmonion for a tetrachord, where the enharmonion has these ratios: $5/4 \times 36/35 \times 28/27 = 4/3$. Archytas seems to have been a scholar who, besides being a skilled mathematician, relied on empirical observation (see [38]); it may well be that he tested the intervals he calculated on a kanon or similar stringed instrument by ear.

The point is that Pythagorean tone mathematics (of which Archytas was the most famous representative in the 4th century) was not confined to the prime numbers 2 and 3. It should be added that Pythagorean tuning in perfect fifths produces a number of nearly just major and minor thirds. If we assume Pythagorean tuning was predominant for medieval organs (as can be concluded from treatises on mensuration of organ pipes and sources relating to the construction of early organs, see [10, 12, 13, 39], a chain of twelve pure fifth (e.g., from a_b to $c^{\#}$) would produce the following scale with c taken as the centre (1/1):

c	$c^{\#}$	d	e_b	e	f	$f^{\#}$	g	a_b	a	b_b	b	c′
0	114	204	294	408	498	612	702	792	906	996	1110	1200

In this tuning (given in modern cents rounded to whole numbers) the major thirds $c^{\#}$–f, e–a_b, $f^{\#}$–b_b and b–e_b are almost just at 384 cents; likewise, the minor thirds e_b–$f^{\#}$, a_b–b and b_b–$c^{\#}$ are almost just at 318 cents. If one wants to avoid accidentals for most of the just intervals, an appropriate segment of the chain of fifths has to be selected accordingly (the chain of fifths can be used like a ‘sliding rule’, see [14]). In a Pythagorean tuning based on the scale as indicated above, the major triads e–a_b–b, b–e_b–$f^{\#}$ and $f^{\#}$–b_b–$c^{\#}$ would have perfect fifths and almost just major and minor thirds. Likewise, the minor triads e_b–$f^{\#}$–b_b, a_b–b–e_b and b_b–$c^{\#}$–f have perfect fifths and nearly just minor and major thirds. The problematic triad in major would be the triad $c^{\#}$–f–a_b which offers a nearly just major third but a narrow

fifth (at 678 cents), and the corresponding minor triad $c^{\#}$–e–a_b which has the nearly just major third e–a_b but the same narrow fifth $c^{\#}$–a_b.

Though Pythagorean tuning perhaps was suited to late medieval organs still conceived as a so-called Blockwerk (several pipe ranks per scale tone of the keyboard mounted on one undivided wind chest, see [39]), it fell short of providing just intervals needed in musical genres that exposed simultaneous thirds. While the use of just major thirds was apparently common in singing (as several theorists assert), a clear indication for a scale different from Pythagorean lore is found in Ramis de Pareia's *Musica practica* (1482/1901). Ramis de Pareia ([40], part I, Chap. 2) gives a division of the monochord that leads to a scale spanning two octaves. Taking a = 1/1 as the tone corresponding to the full string, a scale a–a″ results

a	b	c′	d′	e′	f′	g′	a′	b′	c″	d″	e″	f″	g″	a″
1/1	8/9	5/6	3/4	2/3	5/8	5/9	1/2	15/32	5/12	3/8	1/3	5/16	5/18	1/4

Ramis de Pareia (40, part I, Chap. 5) expands this diatonic scale to a chromatic one, which represents the following segment of a tone-net (cf. [10, 161]):

		d	a	e	b	$f^{\#}$	$c^{\#}$
a_b	e_b	b_b	f	c	g		

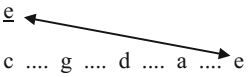
Taking c as the centre, the intervals for the scale would be in modern cents:

c	$c^{\#}$	d	e_b	e	f	$f^{\#}$	g	a_b	a	b_b	b	c′
0	92	182	294	386	498	590	702	792	884	996	1088	1200

This scale has the advantage of including, besides perfect fifths and fourths, four just major thirds, three just minor thirds as well as just major and minor sixths. There are still some Pythagorean intervals (e.g., the minor third c–e_b, the minor sixth c–a_b, the major third e_b–g), however, the just major and minor thirds that could be used for several just major and minor chords (B_b-major, F-major, C-major, d-minor, a-minor, e-minor) would be the main achievement if Ramis' chromatic scale would have been implemented on a keyboard (Ramis addresses the issue of actually tuning his scales in part III, Chaps. 13 and 14 of his treatise). By about 1500, the just major third 5/4 was accepted as a consonance in works on music theory and was used in musical composition. Major chords can be found, for example, ending musical settings assembled in the *Buxheimer Orgelbuch* (ca. 1460-70). The just major third 5/4 has the just minor third 6/5 as a complementary interval within the perfect fifth (5/4 * 6/5 = 3/2), and the two just thirds have the just major sixth (5/3) and the just minor sixth (8/5) as complementary intervals

within the octave (5/3 * 6/5 = 2/1; 8/5 * 5/4 = 2/1). Hence, the quest for just major thirds almost automatically involved tuning several intervals to just intonation ratios.

There are musical works from the 15th century onwards proving that major thirds gained importance in keyboard music. The change from textures based on perfect fifths (as are found in the *Estampie* and the *Retrove* of the Robertsbridge manuscript from c. 1325) to a much more frequent use of thirds as in settings of songs (e.g., *Mit ganczem Willen*) in Conrad Paumann's *Fundamentum organisandi* (Nuremberg/Munich 1452) is obvious, and is continued in organ pieces where major thirds are prominent as in *In dulci jubilo*, contained in Fridolin Sicher's organ tabulature (St. Gallen, c. 1512), or in Hans Kotter's Präambulum in fa (tablature, ca. 1520). Tuning organs in major thirds also must have been explored since Schlick [41], an organist experienced in tuning, writes in his treatise on organ builders and organs (1511) that three just major thirds, stacked upon each other, would be good in quality as such, however, would fail to give one octave as the third tone would be too low (in fact, missing the octave by a diesis of 41 cents). Schlick gave a description of a practical tuning process which would result in a temperament similar to what became known later as 1/4-comma meantone temperament. His tuning aims at just major thirds by slightly narrowing the fifths. Basically, tuning four fifths which are somewhat smaller (in regard to fundamental frequencies) than the ratio 3/2 would yield a tone that is close to a ratio 5/4 relative to the first tone, like, for example,



Taking the difference between the fourth fifth e and the just major third $\underline{e}$ (e^{-1}), which is the syntonic comma of 21.5 cents, it has been equally distributed to the four fifths which are thus narrowed each by c. 5.5 to c. 696.5 cents. The tone d would be the mean (193 cents) between c and e. In a tentative reconstruction of Schlick's meantone temperament [42, 26–29], the scale he tuned would be close to these cents:

c	$c^{\#}$	d	e_b	e	f	$f^{\#}$	g	$g^{\#}$	a	b_b	b	c′
0	76	193	310	386	503	579	696.6	793	890	1007	1083	1200

This temperament offers no less than seven just major thirds and works fairly well for a number of major and minor chords (C, D, E_b, F, G, A, B_b-major; a, b, c, d, e, $f^{\#}$, g-minor) which are in the center of harmonic keys in use at that time. There are some intervals which are problematic in regard to roughness (e.g., the thirds $c^{\#}$–f, $f^{\#}$–b_b, b–e_b are c. 427 cents wide, the fifth $c^{\#}$–$g^{\#}$ has c. 717 cents).

Indications for a temperament that features major thirds and accepts narrowed fifths are found in various sources after 1500 (cf. [8, 12]). However, the tuning instructions that appear as an appendix to the important organ tablature of Johannes of Lublin (Jan z Lublina, c. 1540) still feature perfect fifths with only the two fifths

f–c and g–d narrowed, and several major thirds clearly sharpened; the two fifths have been tentatively estimated at being narrowed by 1/3 of a syntonic comma or ca. 7 cents (see [43]). One reason to keep closer to a Pythagorean type of tuning perhaps was the system of authentic and plagal modes (as elaborated in [44]) and the modal structure in particular of church music; many sources indicate a rather gradual development from medieval psalmody and modal scale concepts to modern tonality as is evident also in secular works for organ (see [45]).

In a treatise of Pietro Aaron (or Aron) the issue of temperament (labelled 'participatione') is addressed where the major third c–e shall be tuned sonorous and just (*sonora et giusta*, [46], cap. XLI). Though *giusta* could be taken to mean 'correct' as well as 'just', *sonora* suggests this major third should be in just frequency ratio (or very nearly so) in order to avoid beats and roughness (as one will experience with Pythagorean major thirds 81/64). If the just major third had become the decisive interval in regard to tuning, perception, and composition of musical works, the system that could provide for a maximum of eight just major thirds contained in a scale of but twelve tones is what we know as 1/4-comma meantone 'temperament'; the term 'temperament' is not quite correct since there are eight just major thirds at the core of the system (while the technical term 'meantone' is from the 19th century and reflects the division of the major third in two equal whole tones). Eight just major thirds are at hand if four pairs of just thirds (b_b–d–$f^{\#}$, f–a–$c^{\#}$, c–e–$g^{\#}$, e_b–g–b) are tuned like

Scheme of quarter-comma meantone temperament for 12 keys

-2		$f^{\#}$ ….	$c^{\#}$ ….	$g^{\#}$	
		|	|	|	
-1		d ….	a ….	e ….	b
		|	|	|	|
0		b_b ….	f ….	c ….	g
					|
+1					e_b
	+11	+5.5	0	-5.5	cent deviation from pure fifth

The just major thirds are in the vertical in this lattice and connected by the sign |. The fifths narrowed by one or two quarters of a syntonic comma are in horizontal direction and connected by…. in this scheme. In 1/4-comma meantone tuning the most problematic interval is $g^{\#}$–e_b, which has 738.5 cents and can hardly be used as a fifth. This was the prize to be paid for the sweetness of so many just thirds and sixths. The scale corresponding to the scheme shown above is

c	$c^{\#}$	d	e_b	e	f	$f^{\#}$	g	$g^{\#}$	a	b_b	b	c′
0	75.5	193	310.5	386	503.5	579	696.5	772	889.5	1007	1082.5	1200

In this meantone scale, d is right in the middle between c and e (consequent to dividing the major third by half). In a straightforward mathematical treatment, the division of the major third can be done like $\sqrt{\frac{5}{4}}$, which yields 1.11803 = 193.2 cents. This, however, is a modern way of calculation that was not feasible by about 1520. In retrospect, the division of the rational interval of the just major third 5/4 into two whole tones 9/8 and 10/9 (which are also rational superparticular intervals, see [10]) as anticipated in the tetrachord divisions of Didymos and Ptolemy (see above) seems of importance since it permitted to build a just diatonic scale suited to intonation of harmonic major.

	9/8	10/9	16/15	9/8	10/9	9/8	16/15	
c		d	e	f	g	a	b	c’
1/1		9/8	5/4	4/3	3/2	5/3	15/8	2/1

Arranged as a small segment of the tone-net, the tones will form a structure like

-1	[d]	a	e	b	
0		f	c	g	d

The tones of this diatonic scale suffice to create three major chords (C, F, G) forming a harmonic cadence. There are also three minor triads (a, e, and d). However, for the minor triad d–a–f, the tone d (in brackets) is not available from this scale. The fifth d–a (of the ratio 40/27 = 680.5 cents) included in the diatonic scale is narrowed by one comma.

Probably the first theorist who understood the dilemma of tuning just fifths and major thirds in regard to building a chromatic scale with a rather small number of scale steps was Fogliano. He (1529, fol. xxxv) presented a chromatic scale which doubles the tones d and b_b; as he uses both the diatonic (16/15) and the chromatic (25/24) semitone as well as the syntonic comma 81/80, the following scale results:

c	$c^{\#}$	d	d	e_b	e	f	$f^{\#}$	g	$g^{\#}$	a	b_b	$\overline{b}_b$	b	c′
1/1	25/24	10/9	9/8	6/5	5/4	4/3	25/18	3/2	25/16	5/3	16/9	9/5	15/8	2/1

Ordered into a tone-net, the fourteen tones per octave result in this structure:

-2	$f^{\#}$	$c^{\#}$	$g^{\#}$		
-1	d	a	e	b	
0	b_b	f	c	g	d
+1				e_b	b_b

The expansion of the chromatic scale from twelve to but fourteen tones per octave yields no less than eight just major chords and seven just minor chords. In addition, it offers chromatic and diatonic semitones (for example, $\underline{e}$–e_b, c–$c^{\#}$, f–$f^{\#}$, g–$g^{\#}$ = 25/24 = 70.7 cents; $\underline{d}$–$c^{\#}$, $\underline{e}$–f, b_b–$\underline{a}$ = 16/15 = 111.7 cents) as well as a semitone 135/128 = 92.2 cents (the step b_b–$\underline{b}$) and a semitone 27/25 = 133.2 cents ($f^{\#-2}$–g). In addition, there is the fourth $\underline{d}$–g and the fourth f–$\overline{b}_b$ which have a ratio of 27/20 = 519.6 cents. The cost for realizing Fogliano's scale in a keyboard instrument would be adding two extra keys and strings or pipes per octave. Fogliano did not see this as practical and considered tuning a tone halfway between the two doubled tones (d, d^{-1} and b_b, b_b^{+1}) instead. In a modern approach, the geometric mean of the whole tones 10/9 and 9/8 would be calculated like $(\sqrt{5})/2 = 1.11803$, which equals 193.2 cents, the size of the meantone. Though Fogliano's 14-tone scale apparently was a construct devised to solve a problem in music theory, it was well within the possibilities of instrument building in the 16th and 17th centuries when indeed a considerable number of organs and harpsichords had one or several split keys (see [1, 3, 12, 47]).

Meantone tunings (of which several varieties were in use) featuring just major thirds basically face the same problem one experiences with Pythagorean chains of perfect fifths: they do not easily lead to a cycle within an octave that comprises no more than twelve tones and keys. Therefore, adding at least one or two tones and keys per octave was inevitable if the so-called "wolf" $g^{\#}$–e_b was to be eliminated. The most common solutions were split keys for $d^{\#}/e_b$ and $g^{\#}/a_b$ as well as, in a smaller number of instruments, $a^{\#}/b_b$. Expanding meantone tuning even further, split keys for all accidentals were implemented (=17 tones/keys per octave). Adding two more tones and keys for $e^{\#}/f_b$ and $b^{\#}/c_b$ results in a 19-tone cembalo cromatico (as shown in [48, 141]) and described, for the cembalo universale owned by Karel Luyton at Prague, in Praetorius' treatise ([49, T. II, 63ff.]) on music and musical instruments.

Assuming that a maximum of just major thirds was the main purpose for developing extended meantone tunings, the scheme of 19 tones implemented in the cembalo cromatico and, by comparison, the 24 tones/keys per octave Zarlino had on his enharmonic harpsichord built in 1548 (cf. [3, 17ff.]) can be shown in a tone lattice like

+4					$g^{\#\#}$	$d^{\#\#}$	$a^{\#\#}$	
+3	$e^{\#}$	$b^{\#}$			$e^{\#}$	$b^{\#}$	$f^{\#\#}$	$c^{\#\#}$
+2	$c^{\#}$	$g^{\#}$	$d^{\#}$	$a^{\#}$	$c^{\#}$	$g^{\#}$	$d^{\#}$	$a^{\#}$
+1	a	e	b	$f^{\#}$	a	e	b	$f^{\#}$
0	f	c	g	d	f	c	g	d
-1	d_b	a_b	e_b	b_b	d_b	a_b	e_b	b_b
-2				g_b				g_b

For an arrangement of keys and pitches shown in the tone lattice on the left, the term 'cembalo cromatico' seems not quite correct since there are intervals smaller than the chromatic semitone 25/24 (70.7 cents). Hence, enharmonic melodic phrases could be realized making use of the difference in pitch between the sharps and the flats (e.g., d_b–$c^{\#}$ etc.) as well as between $b^{\#}$ and c, $e^{\#}$ and f. Zarlino's instrument in fact was suited to playing enharmonic intervals and melodic phrases. Vicentino [50] had the number of tones and pitches on his keyboard expanded to 31 to the octave (a similar instrument was built, in 1606, by Vitus de Trasuntino; for his 'Clavemusicum omnitonum', see [3, 25f.]).

Zarlino, in Part II of the *Istitutioni*, offers an in-depth elaboration of tetrachord divisions, scales and aspects of tuning strings on a monochord in which he refers to Greek writers, in particular to Ptolemy and his tetrachord divisions; the diatonon syntonon was of special interest to Zarlino because of the just major third and its division into two whole tones of different ratio and interval size. In regard to tuning keyed instruments (such as a gravecembalo), Zarlino distinguishes between 'natural' intervals and temperaments (*temperamento o participatione*). After briefly mentioning his own instrument (a *clavocembalo* built, in 1548, by Domenico da Pesare; see Istituioni 1558, 140/41) suited to realize chromatic and enharmonic harmony, Zarlino [51] refers to yet another instrument of which more would be said in his *Demostrationi harmoniche* (published in 1571 but apparently written at the same time as his first book). In this book, however, there is again only a brief passage (on p. "212", which is the wrongly numbered p. 221) while a full description is found in Zarlino's *Sopplimenti musicali* (1588, cap. XI). Zarlino discusses a tone system (systema massimo arteficiale del naturale ò syntono diatonico) which comprises 33 tones and pitches in a two-octave range (A–a–aa). He takes whole numbers (a method known since Aristides Quintilianus and Boethius, see [10, 147] which can be taken to represent the distances between tones on a string of a monochord. Transferred into modern cents, the structure of his scale may be shown here only for the lower octave:

A 112 B_b 21.5 B_b 71 B 112 c 70 $c^{\#}$ 113 d 21.5 d 90 e_b 21.5 e_b 71 e 112 f 71 $f^{\#}$ 21.5 $f^{\#}$ 112 g 71 $g^{\#}$ 112 a

Relative to c, the lattice of tones would contain

-2		$f^{\#}$	$c^{\#}$	$g^{\#}$		
-1		d	a	e	b	$f^{\#}$
0	e_b	b_b	f	c	g	d
+1					e_b	b_b

Implemented on a keyboard (as shown by [52, 156]), these 16 tones and pitches per octave would offer a range of eight just major and seven minor triads. It seems the solution proposed by Fogliano [53] in regard to a doubled d and a doubled b_b thereby was used for the tuning of an advanced keyboard instrument (though the doubled e_b in Zarlino's scale hardly offers any benefit – while a $d^{\# -2}$ or an a_{b+1} tuned instead would have).

Though Zarlino discusses Greek scales in his treatises extensively, it was not his intention to revive the music of the 'antichi' in the sense of using chromatic or even enharmonic scale models based on tetrachords. Rather, his goal was to explore intervals and chords in regard to just intonation as is obvious from his own compositions, in particular the *Modulationes sex vocum* [54], a collection of motets published in 1566 (a critical edition by Collins Judd and Schiltz was published in 2015; a recording by the ensemble Singer Pur of Munich was issued in 2013). These works are in the tradition of Adrian Willaert and the vocal polyphony for which Venice was famous. Zarlino composed motets rich in harmony based on the just thirds he had justified, in the *Istitutioni*, with his concept of the *senario*. In this respect, his approach was different from that of some contemporaries, among them Nicola Vicentino, who apparently had a more experimental attitude towards the use of chromatic and enharmonic intervals in musical settings. A good example is the small madrigal ('madrigaletto') *Dolce mio ben*, of which Vicentino (1555, cap. LII) offers three versions, one in the diatonic genus, one in the chromatic, and one in the enharmonic (for a detailed analysis including sound examples of the different versions, see Cordes [33]). For the rehearsals with his students as well as for demonstrations, Vicentino used an archicembalo that had 31 tones to the octave. The exact tuning of the instrument has been a matter of debate since the two schemes Vicentino offers for tuning allow for some interpretation (see [12, 390ff.]). However, a model where just thirds are piled up in the vertical in six or seven rows (cf. [3, 25]) seems plausible since it can be taken as a further extension of the meantone tuning beyond the 19-tone cembalo cromatico (see above). If Vicentino's division of the whole tone into five parts is taken as meaning interval steps of equal size, a regular temperament could be assumed where the diesis of 38.71 cents is the basic unit. Multiples of this unit result in the chromatic and in the diatonic semitone

(77.5 and 116.1 cents, respectively), the minor and the major third (at 310 and 387 cents, respectively), and the minor and major sixth (at 813 and 890 cents, respectively). This system, described much later (1661/1691) with mathematical background by Christiaan Huygens (see [6, 7]), offers a range of nearly just intervals (including a 'natural' seventh close to the ratio 7/4 at 968 cents) but maintains the slightly narrowed fifths and slightly widened fourths as well as the whole tone (at 193.55 cents) halfway between 10/9 and 9/8. In this respect, Huygens' cycle is an expansion of the meantone system (with a number of additional tones and pitches that were of little use in Baroque music but became a means for contemporary music in the 20th century, see [7]). Implementing 31 tones and pitches to the octave on a keyboard is a demanding task for both the instrument builder (skilfully mastered by Trasuntino and other artisans) and the musician who must adapt to a keyboard with at least three rows of keys. Vicentino was not the only enharmonic experimentalist. There were more instruments with more than 19 keys to the octave in use (see [3, 47]). A late specimen of a sophisticated keyboard with 31 keys to the octave is a Hammerclavier built by Johann Jakob Könnicke, in 1796 (see the photo in [3, 465]). The keys are ordered in a very intelligent fashion, which makes playing certain chord patterns fairly easy (a description of the arrangement of keys and pitches is given by Vogel [14], 304-08 and pp. 319-23 in the English edition of 1993). The 31-tone pipe organ which was built in the Netherlands in 1945 also offers a special keyboard designed by Fokker (see photos in [7]) which permits to play sequences of major or minor chords by shifting the hands in diagonals without changing the fingering.

Of course, raising the number of pitches and keys per octave in a regular division improves the approximations to just intonation pitches. While a division of the octave into 31 dieses of 38.71 cents each is sufficient to produce nearly just thirds and sixths as well as the 'natural' seventh, just fifths and fourths require a division of the octave into 53 equal parts of 22.64 cents each. Evidently, the unit here is a 'comma' (close in size to the syntonic comma of 21.5 cents), the multiples of which will give suited musical intervals (e.g., the sum of 17 commas yields a major third of 385 cents, 22 commas make up a perfect fourth of 498 cents, the sum of 31 commas gives a perfect fifth of 702 cents, etc.). The division of the octave into 53 equal steps, which seems to have been calculated by the mathematician Nicolaus Mercator by about 1660 (he first calculated a 'comma' corresponding to a division of the octave into 55 equal parts), found renewed interest in the 19th century (see [3]). There are more such equal divisions (e.g., 72 pitches and scale steps to the octave), some of which have been used in composition and in the performance of microtonal works by making use of electronic keyboard instruments (see [55]). However, in a historical perspective, mechanical instruments were difficult (and costly) to build with more than 12–14 keys per octave. Even though, the chromatic and enharmonic keyboard instruments that were built, in particular in Italy in the 16th and 17th centuries (for a survey, see [2, 3, 47]), respectively, greatly supported musical practice which saw a range of highly chromatic works for keyboards written by, among others, Merulo, Mayone, and M. Rossi.

Even the 'standard' 1/4-comma meantone tuning confined to 12 keys and pitches to the octave supports chromatic expression to some degree since it offers a diatonic (117.5 cents) and a chromatic semitone (75.5 cents) which are audibly distinct. Progressions in semitones as are found frequently in keyboard works of the 17th century (written by, among others, Sweelinck, Bull, Philips, Schildt, Froberger), when played on harpsichords and organs tuned to 1/4-comma or one of the meantone varieties, are of interest to listeners who may recognize different interval sizes. Keys available in a common meantone tuning with good sound quality typically span from E_b-major to A-major (that is, from three flats to three sharps). There are works in E-major like the Praeludium in E from Dietrich Buxtehude (BuxW 141) which can also be played on an organ in meantone tuning (with cautious registration in regard to the use of mixture stops and still accepting a few relatively harsh sonorities), and even many of Bach's organ works can be played on an organ tuned to 1/4-comma meantone though there are some parts in a number of works that sound quite harsh in this temperament (cf. [18]). A scale of but 12 pitches to the octave for a number of Bach's organ and harpsichord works seems insufficient since, for example, in the Fantasie und Fuge in g-minor (BWV 542), for the harmonic modulation found in measures 31–38 of the Fantasie, one would need a total of about 25 different pitches and tones if this part would be played in just intonation, that is, with perfect fifths and fourths as well as with just major and minor thirds. Of course, many works for keyboards of the 17th and early 18th century were far less bold in their harmonic structure, and restricted to those keys and chords which turn out to be pleasing in their sound in meantone tuning. To be sure, the meantone concept was developed with the major third as the basic structural interval in mind, and in regard to the 'sweetness' of simultaneous thirds and sixths it could offer to the player and listener alike. It was for this effect that various composers adapted Dowland's *Lachrimae* to versions for keyboard instruments.

One has to remember that compositional practice in the 17th century and even in the first half of the 18th century still included the regular use of modal scales and melodic patterns while chord progressions were formed in simple or extended cadences that established the concept of major and minor tonalities, respectively (elements fundamental to this new concept were discussed, for example by Rameau in his books on music theory of the 1720s and 1730s, see [56, 57]). A harmonic tonality typically involves a centre expressed by a major or a minor chord in a certain key from which one can modulate into adjacent or more distant keys. The 'distance' thereby in general is conceived in terms of fifths, and the geometric structure to represent keys is known as the 'cycle of fifths'. In 'western' music theory, ideas on such a cycle were issued before 1700 (for example, by A. Kircher). A more formal discussion on the relationship of tones was offered by Johann Heinichen who, in 1711, published a 'musical cycle' that shows the tones and keys actually used at his time plus a few more distant tones and keys that were conceivable yet not practical. A revised version of the Musicalischer Circul was published by Heinichen [58]. Heinichen ([59, 261ff.]) explains that the use of tones and keys in practice could go as far as B-major on the side of the sharps around the

circle, and to b_b-minor on the side of the flats, both taken as 'extremes'. This would mean twenty out of twenty-four major and minor chords and keys were in use. In Johann Fischer's *Ariadne Musica* (1702, 1710) there are twenty tonalities, ten major (A_b, E_b, B_b, F, C, G, D, A, E, B), nine minor (f, c, g, d, a, e, b, $f^{\#}$, $c^{\#}$), and e-Phrygian. However, the e-minor is conceived as e-'Dorian', and there are more modal remnants in Fischer's cycle (see [45]). Heinichen [59] warned that the use of the most distant keys and chords in his cycle would be of no avail. One possible interpretation of his statement could be that these distant keys are too remote in regard to forming meaningful sequences of keys and chord progressions relative to a well-established tonal centre (which he identifies as C-major). Another aspect possibly included in Heinichen's discussion is that of tunings and temperaments. Though it is relatively certain that 1/4-comma meantone tuning remained the standard in many areas of Europe well into the 18th century (see [1]), it is also known from various sources that organ builders and organists experimented with temperaments where the size of fifths and thirds varied in such a way that certain keys were quite smooth in regard to roughness and beats (the 'good keys') while others were more harsh in particular when chords were played with a registration that involved mixture stops (which to this day are tuned in just intervals). In the period from c. 1680 to c. 1770 various 'well-tempered' tunings were proposed and/or explored in practice (see [8, Chap. 7]). Werckmeister offered several tunings of which Werckmeister III (sometimes also counted as no. IV) became well-known as "the" Werckmeister tuning model (see [19, 60]). The concept of this tuning was a closed circle of fifths, which means that several or all fifths need to be narrowed in order to distribute the 'overshoot' (see above) of a Pythagorean comma (ca. 24 cents). In Werckmeister III there are four fifths (c–g, g–d, d–a, and b–$f^{\#}$) which are narrowed by a quarter of the Pythagorean comma ([4, 161]; [60]). The following scale results (rounded to full cents):

c	$c^{\#}$	d	e_b	e	f	$f^{\#}$	g	$g^{\#}$	a	b_b	b	c′
0	90	192	294	390	498	588	696	792	888	996	1092	1200

In this scale, the d is still a meantone, c–e_b comes as a Pythagorean minor third, and the fifth c–g is of nearly the same size as the tempered fifth in 1/4-comma meantone whereas the fourth c–f here is perfect, and the b_b is slightly flattened and the b sharpened in comparison to 1/4-comma meantone. The major third is still quite good though a C-major chord suffers from the third being slightly too wide and the fifth being narrowed, the interval between them a minor third of 306 cents. The third f–a (390 cents) is good and g–b (396 cents) acceptable, however, the thirds $c^{\#}$–f, $f^{\#}$–b_b and $g^{\#}$–c are Pythagorean (408 cents). The major thirds e_b–g, e–$g^{\#}$ and a–$c^{\#}$ all have 402 cents. In Werckmeister III chords in the center (C-major, F-major, G-major, D-major) appear quite fair relative to just intonation intervals while triads in keys with more accidentals are less satisfactory. In this respect, major and minor chords in various keys can be distinguished by their sonorous

quality (for data, see below) while there is no obvious discordance in Werckmeister III like the 'wolf' in 1/4-comma meantone. Thus, Werckmeister III would support modulation through a wider range of keys as is suggested by Heinichen [58, 59]. It is a common feature of 'well-tempered' tuning models discussed or empirically tested that they seek to allow modulation through most or even all (commonly accepted) major and minor keys while maintaining some musical and perceptual discriminability between different keys.

The important achievement of the 1/4-comma meantone tuning had been a maximum of eight major thirds out of a scale comprising, in its basic form, only twelve tones and pitches to the octave, at the cost of the 'wolf fifth' as well as some other relatively poor intervals. The 'well-tempered' tunings could remedy the obvious defects of 1/4-comma meantone yet had to sacrifice the just major thirds to some extent. In sum, one can see that the improvement of the fifths in 'well-tempered' systems as well as the possibility for harmonic modulation through many keys was kind of an intermediate solution between the Pythagorean approach (just fifths and fourths plus a few nearly just thirds) and the meantone concept (numerous just thirds and sixths, tempered fifths and fourths). 'Well-tempering' in many instances was derived from the experience of tuning keyboards as apparently was the case with J.S. Bach who tuned his own instruments (there are legions of interpretations what 'well-tempered' may have been for Bach and his 'Well-Tempered Clavier', see [8, 15–17, 61]). Some of the more theoretical approaches (e.g. [19]) to finding the 'very best temperament' still made use of geometrical tools such as dividing strings on a monochord into sections, or tried to calculate equal temperaments from a basically geometric perspective (as did Neidhardt in a number of studies, see [8, 264ff.]). Of course, there were also attempts at finding a circular equal temperament in an algebraic calculation. The means for such calculations included logarithms which had been developed already in the 16th century. However, sources indicate that Juan Caramuel Lobkowitz in about 1647 was the first to suggest logarithms to base 2 as a measure suited to calculate and represent musical intervals (see [3, 282ff.]). The mathematicians Isaac Newton and Leonhard Euler also contributed to such calculations. In the 19th century, another measure was proposed by the French acoustician, Felix Savart, which defines 1 octave = 1000 $\log_2$ (=301.03 Savart, see [62, 3f.]). Further, the physicist Arthur von Oettingen calculated intervals as milli-octaves, mo (cf. [14, 111]). The mo, which is 1/1000 of an octave, can be expressed like

$$1\,\text{mo} = \sqrt[1000]{2} = 2^{1/1000} = 1.000934.$$

Thus, a sine tone differing from a standard (say, A_4 = 440 Hz) by 1 mo, would have a frequency of 440 × 1.000934 = 440.4109 Hz. The pure fifth (3/2) has 585 mo, the just major third has 322 mo, the just minor third 263 mo. The advantage of the mo is that just intervals result in whole numbers.

Since the octave typically comprises twelve semi-tones (of equal or unequal size), a division of the octave into 1200 basic units rather than 1000 mo seemed appropriate. Alexander J. Ellis suggested the modern cent as $\sqrt[1200]{2} = 2^{1/1200}$,

whereby 1 cent = 1.00058. This unit is convenient for expressing musical intervals in ET12 where 1 semitone = 100 cents, meaning all intervals in ET12 are multiples of 100 cent. However, their frequency ratios are complicated consequent to the tempering which, in ET12, defines the fifth as 1:1.498307 (=700 cents) and the frequency ratio of the major third as 1:1.259921 (=400 cents). Representing two sine tones (in this article, a tone is considered as a physical phenomenon notwithstanding its musical functions) each by a single frequency, f_1 and f_2, the interval they form can be expressed as the ratio f_2:f_1 and the interval can be calculated in cents like $1200 \log_2 (f_2/f_1)$. For example, taking two sine tones of 200 and 300 Hz, respectively, the pure (or just) fifth thereby can be calculated like

$$1200 \log_2(300/200) = 701.955 \text{ cents}.$$

The difference between structurally important intervals in just intonation and ET12 is this:

Interval	Just	ET12
Fifth	702	700
Fourth	498	500
Major third	386	400
Minor third	316	300
Major sixth	884	900
Minor sixth	814	800
Minor seventh	969	1000

The largest deviation from a just interval thus is about 16 cents, with the exception of the minor seventh. If one accepts that, for example in a dominant seventh chord, the seventh should be of the ratio 7/4 (see [14]), corresponding to the 'natural seventh' (the seventh harmonic in a harmonic partial structure), the deviation in ET12 from the just interval is more than 31 cents.

The quest for ET12 can be viewed as a solution to the prime number discrepancy stated as $3^n \neq 2^m$ and $5^n \neq 2^m$. In order to derive cyclic scales closed within each octave, some adjustment of the size of intervals is necessary (cf. [11]). This led to concepts of regular as well as irregular temperaments (meanings of the Latin noun *temperamentum* include 'the right measure'). A regular temperament does not imply that all scale steps are of the same size (see [8]). However, a regular temperament can be established by dividing the octave into k equal parts. With a division into twelve parts, ET12 can be realized as a tuning (the term tuning rather denotes the actual process of pitch adjustment than the calculation of pitch frequencies or pitch ratios). In ET12, the deviations from just intonation are small for fifths and fourths yet considerable for thirds and sixths, putting ET12 relatively close to Pythagorean tuning. If one prefers a temperament and tuning that offers nearly just thirds and sixths as well as 'natural sevenths', ET31 would be the

choice. In case fifths and fourths as well as thirds and sixths should be close to just intonation pitches, ET53 seems the best solution. If the prime number 7 is taken into account in addition to the prime numbers 2, 3 and 5 for the generation of musical intervals, a division of the octave into 171 small steps gives the best approximation to just pitch ratios (see [14]).

The choice for a particular regular or irregular temperament may be guided by certain criteria such as the maximum deviation from just intervals one is willing to accept in terms of cents (for mathematical models and calculations of scale models and tunings, see e.g. [63, 11, 17, 9]), or the amount of roughness and beats one may allow in simultaneous intervals and chords (see [29]). ET12 can be regarded a good compromise since it offers (1) a closed cycle of tones per octave as well as (2) usability of twelve major and twelve minor keys. In the 17th and well into the 18th century, exact calculation of ET12 pitch ratios was a problem, and actually tuning an organ to ET12 was difficult because ET12 involves irrational pitch ratios, on the one hand, and quite irregular beat frequencies, on the other (the German term *gleichschwebende Temperatur* for ET12 is misleading. While the size of semitones in ET12 is fixed, beat frequencies vary for the eleven intervals within different octaves).

The mathematical solution for ET12 nowadays is straightforward by solving the equation (in the syntax of Mathematica©)

Solve $[x^{12} = 2^7, x]//\mathrm{N}$ or solving the equation Solve $[(x/2)\text{^}12 - (2/1)\text{^}7 == 0, x]//\mathrm{N}$

For x, a set of solutions is obtained which includes $x \rightarrow 1.49831$, meaning the size of the fifths must be narrowed from a ratio of 3:2 (or 1.5:1) to 1.49831:1 to make 12 fifths equal 7 octaves. In fact, the number 1.49831 indicates an interval size of 700 cents (the fifths in ET12) which results from distributing the overshoot of 24 cents equally to 12 fifths. Likewise, the frequency ratio for the semitone in ET12 can be found from the equation

$$\text{Solve}\,[2^7 + x^{12} == 0, x]//\mathrm{N}$$

where the set of solutions includes $x \rightarrow 1.05946$, which equals $\sqrt[12]{2}$ = 100 cents. From here, finding the major third in ET12 is easy since $\left(\sqrt[12]{2}\right)^4$ = 1.25991 ≈ 400 cents.

Though ET12 appears as an elegant solution in that it distributes the Pythagorean comma equally to twelve fifths, it met considerable resistance in the 18th century because it practically eliminated musical and perceptual differences between keys. As an alternative, various temperaments were explored which affect the basic intervals (semitones, tones, fifths, fourths, thirds, sixths) to different degrees (see [8]). Solutions depend on decisions one makes in order to keeping certain intervals close to just frequency ratios while others then will deviate a bit more from just ratios. Such decisions in general have effects on the musical keys and chord textures that sound smoothly within a given temperature and tuning. Deviations from just intervals must be small enough to avoid whatever perceptions of mistuning of

certain scale steps and intervals. Among the temperaments that met this requirement is 1/6-comma meantone, where the fifths are narrowed by 1/6 of a syntonic comma (3.6 cents, see [64, 456]) to c. 698 cents, to the effect that most of the major thirds (c–e, e–$g^{\#}$, f–a, a–$c^{\#}$, g–b, d–$f^{\#}$, e_b–g, b_b–d) are close to 393 cents. The 'wolf' between $g^{\#}$ and e_b is not eliminated but is reduced to 718 cents. There are three rather problematic major thirds ($f^{\#}$–b_b, $c^{\#}$–f, $g^{\#}$–c) which have c. 413 cents, and, correspondingly, there are three problematic minor thirds (b_b–$c^{\#}$, f–$g^{\#}$, e_b–$f^{\#}$) which are significantly narrow.

Among the temperaments that have gained importance is one attributed to the Italian composer, organist and theorist, Francesco Antonio Vallotti (a part of his work was published in 1779, while the part containing his concept of temperament was left in manuscript and published only in 1950; see [8, 306]). A very similar temperament was devised by the English scientist, Thomas Young (who actually proposed two temperaments in 1800). The basic idea in Vallotti's temperament is to tune six fifths f–c–g–d–a–e–h so that each fifth is narrowed by 1/6 of a comma (to 698 cents), and to tune another six fifth in just frequency ratios. Correctly notated, these intervals would be f–b_b–e_b–a_b–d_b–g_b–c_b, however, usually the tones are given as f–b_b–e_b–a_b–d_b–g_b–b or, if tuning in upward direction is chosen, as b–$f^{\#}$–$c^{\#}$–$g^{\#}$–e_b–b_b–f in order to underline the circular character of this temperament. The scale then has these tones and intervals (rounded to full cents):

c	$c^{\#}$	d	e_b	e	f	$f^{\#}$	g	$g^{\#}$	a	b_b	b	c′
0	94	196	298	392	502	592	698	796	894	1000	1090	1200

In this temperament, there are major thirds of different size. Major thirds in the middle of the tonal area (f–a, c–e, g–b) have 392 cents, and b_b–d and d–$f^{\#}$ have 396 cents. The thirds e_b–g and a–$c^{\#}$ have 400 cents, e–$g^{\#}$ and $g^{\#}$–c have 404 cents, and b–e_b, $f^{\#}$–b_b, and $c^{\#}$–f have 408 cents, respectively. Hence, there is a gradation in the thirds from those relatively close to the just ratio to thirds close to ET12, and further on to a few major thirds which equal the Pythagorean ditonos. Correspondingly, there is a number of Pythagorean minor thirds of 294 cents (e.g., e_b–$f^{\#}$, $g^{\#}$–b_b) while the minor thirds closest to just ratios are a–c, e–g, and b–d, each of 306 cents. The remaining minor thirds are in between (see [64, 457]). The obvious advantage of the Vallotti temperament is that no 'wolf' interval is encountered, and that modulation through all major and minor tonalities seems possible, though with increasing deviations from just tuning towards the periphery. The gradation of intervals and chords in regard to roughness and sensory consonance can help to differentiate between keys and tonalities, and may be appreciated by listeners. It is in fact interesting to listen to Beethoven's piano works when performed in temperaments and tunings such as that proposed by Vallotti (and, with small variations, by Thomas Young).

4 Empirical Investigation of Temperaments and Tunings

In a number of studies, deviations of intervals in various temperaments from just ratios have been calculated [8, 11, 17, 63–65]. Some investigations also include calculations made from the scores of musical works where the occurrence of certain intervals and chords has been considered for such calculations. When synthesizers and digital signal processing methodology became available to sound and music research, investigations could be expanded from scores to recordings of music, and different tunings could be studied by manipulating sound parameters (see, e.g., [3, 18]).

The present study makes use of signal processing methodology in that the periodicity and harmonicity of major and minor chords is measured in the time domain using autocorrelation (AC) and crosscorrelation (CC) tools developed by Boersma [66]. These tools measure the harmonics-to-noise ratio (HNR) for a given time signal $x(t)$ which is expressed in dB. The sensitivity of the tools depends on jitter in time signals and hence on the frequency and energy distribution of spectral components as well as on temporal factors. In this respect, the dB readings allow a relative scaling of signals in regard to their periodicity. The maximum that we attained with a perfect major chord composed of three harmonic complexes each comprising ten harmonics locked in zero phase with attenuation of amplitudes like $A_n = 1/n$ was ≥ 60 dB.

In a previous study of 1/4-comma meantone tuned with precision on a historical organ built by Arp Schnitger (Hollern, Northern Germany, 1688), a clear gradation for twelve major and twelve minor chords was found [67]. Concerning major chords, there is a grading from very good (C, D) to good (G, A, E, E_b, F), while B and $G^{\#}$ appear as less acceptable, and $C^{\#}$ and $F^{\#}$ are problematic given their low HNR readings. Likewise, for minor chords, a-minor and d-minor are best, followed by $f^{\#}$-minor and e-minor while c-minor and $c^{\#}$-minor gave low readings.

In this study, data for two 'well-tempered' systems will be presented, namely Werckmeister III and Vallotti as tuned on a harpsichord. For our investigation, Werckmeister III and Vallotti as well as 1/4-comma meantone and some other systems were tuned on a historical Jacob Kirckman harpsichord (London c. 1766) from the collection of the second author (see [68, no. 60, 216–223]). This instrument is of interest for some extraordinary mechanical and acoustical features (see [69]). For the recordings, only one 8' stop was used and the strings of all other stops were dampened with cloth. The recordings were made with a single condenser mic (Neumann TLM 170) placed ca. 40 cm over the strings. The sound was recorded on DAT at 48 kHz/16 bit. The tuning was done relative to A_4 = 408 Hz, which is a common pitch for historical harpsichords. For the tuning, a precision digital device (TLA CTS 5-PE) was used which reads fundamental frequencies of sounds radiated from the instrument. The tuning was checked by means of spectral analysis and f_0 tracking of sounds recorded from single complex harmonic tones.

The point where the actual plucking takes place divides each string of a harpsichord into two parts from where waves propagate into opposite direction. Because

partials which have a node at or near the plucking point cannot be excited to undergo vibration (as was reported by Thomas Young, in 1800), the amplitude spectrum shows characteristic troughs and dips defined by L/l (L = string length, l = plucking point measured from bridge; see [26, 27]). Since certain partials are weak or even cancelled out, the spectrum for the sound from each string becomes more or less cyclic as is shown in Fig. 2 (where a formant filter envelope is included that also shows the peaks and dips in spectral energy distribution).

The dips in spectral energy found in the sound of a single string are levelled out to some extent when several strings are played simultaneously in a chord, and partial frequencies of several tones coincide, as can be expected in particular if several tones of the chord are doubled at the octave as was the case in our experiment. The major and minor chords played on the Kirckman comprised five notes and tones each, for example, C-major consists of c_2, g_2, c_3, e_3, g_3, while for $C^{\#}$-major the notes are simply shifted in parallel by one semitone upward, for D-major by a whole tone, etc. The recordings were actually done twice, one run starting at A (because it serves as referent also for the historical tunings in the electronic tuner we used), the other at C. Because of the large number of partials contained already in the sound of individual strings, the spectrum for each chord is rather dense. Figure 3 shows the spectrum for the $C^{\#}$-major chord where the fundamental $c_2^{\#}$ is at 63.99 Hz, and significant spectral energy is found up to 6 kHz (all amplitudes are given relative to 0 dbfs).

Looking closer into the spectrum of the $C^{\#}$-major chord reveals several partials from different tones of the chord differ slightly in their respective frequency. While coincidence of harmonic partials from tones in a chord since long has been recognized as a factor relevant for sensory consonance [70], small divergence in frequency of such partials (each of them carrying sufficient energy) gives rise to

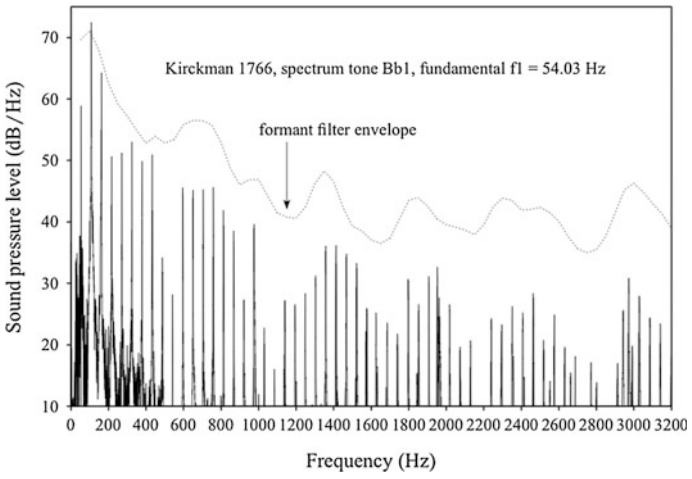


Fig. 2 Kirckman 1766, tone/string Bb_1, sound spectrum, f_1 = 54.03 Hz

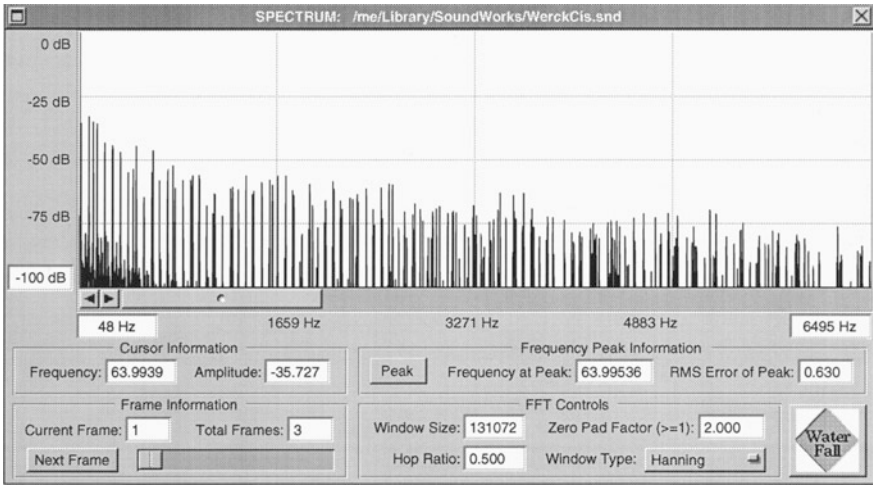


Fig. 3 Spectrum, Kirckman 1766, C$^{\#}$-major chord, Werckmeister III

auditory roughness (see [28, 29]). A clear sign of spectral inharmonicity is amplitude modulation (AM) visible in the temporal envelope of partials (see Figs. 6 and 7) as well as in the envelope of the complex signal representing a chord.

The HNR readings for major and minor chords in Werckmeister III are listed in Table 1. For each chord, decibels represent the means for HNR averaged over two seconds of sound from the onset and the standard deviation (SD) for the same segment. Taking the two first seconds of each chord seems sufficient since, due to the plucking mechanism of strings on a harpsichord, the sound level reaches maximum typically within c. 100–150 ms and then decays smoothly. For the C$^{\#}$-major chord shown in Fig. 4, the decay after two seconds is c. 12 dB from maximum.

Table 1 HNR data, Werckmeister III

Chord	dB (mean)	dB (SD)	Chord	dB (mean)	dB (SD)
C-major	14.87	3.39	c-minor	11.09	2.06
C$^{\#}$-major	10.42	2.65	c$^{\#}$-minor	7.46	1.22
D-major	11.9	2.39	d-minor	11.91	2.38
E$_b$-major	10.61	2.7	e$_b$-minor	6.0	2.65
E-Major	11.92	2.22	e-minor	4.02	1.48
F-Major	18.85	3.53	f-minor	9.01	2.59
F$^{\#}$-major	10.29	2.73	f$^{\#}$-minor	6.59	1.99
G-Major	13.74	2.57	g-minor	7.57	1.45
A$_b$-major	12.25	3.02	a$_b$-minor	8.45	2.22
A-major	15.29	3.77	a-minor	4.85	1.22
B$_b$-major	11.94	2.52	b$_b$-minor	6.63	2.46
B-major	10.14	2.49	b-minor	7.57	1.22

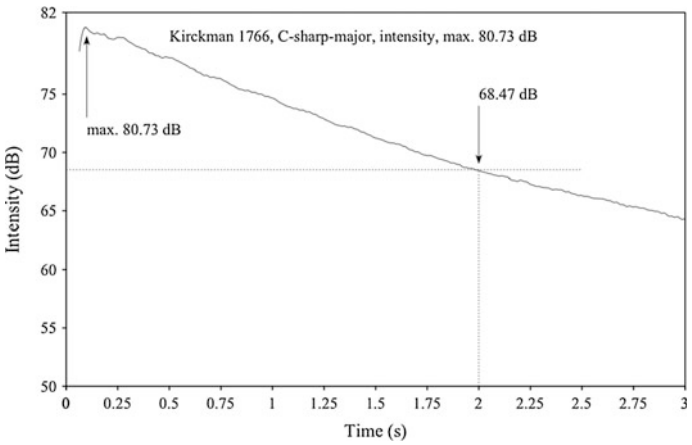


Fig. 4 Kirckman 1766, C$^{\#}$-major, intensity (dB) over time for the first 3 s

The data for the Vallotti tuning are given in Table 2.

An inspection of the data reveals, first of all, a significant difference between HNR for major and for minor chords that are due to differences in their harmonic structure. Such differences were observed also for major and minor chords in 1/4-comma meantone (see [67]). Another factor that seems of interest is the relatively large standard deviation calculated from the HNR data for various major and minor chords. In this context it should be recalled that measurements of sound signals give high readings for HNR with low SD if a signal is strictly periodic in the time domain (Fig. 1), which implies it is strictly harmonic in the spectral domain. Since in both Werckmeister and Vallotti intervals in major and minor chords deviate to some extent from just ratios, interference between pairs or groups of

Table 2 HNR data, Vallotti tuning

Chord	dB (mean)	dB (SD)	Chord	dB (mean)	dB (SD)
C-major	15.99	5.18	c-minor	10.2	1.47
C$^{\#}$-major	11.46	3.42	c$^{\#}$-minor	9.11	2.82
D-major	13.53	3.36	d-minor	3.92	1.65
E$_b$-major	9.8	2.21	e$_b$-minor	3.94	2.27
E-Major	13.23	2.65	e-minor	3.29	0.82
F-Major	14.78	3.29	f-minor	6.32	2.15
F$^{\#}$-major	11.45	3.53	f$^{\#}$-minor	5.32	1.55
G-Major	16.27	4.53	g-minor	7.75	0.94
A$_b$-major	12.19	2.65	a$_b$-minor	9.05	2.79
A-major	14.24	3.4	a-minor	11.25	2.32
B$_b$-major	16.17	3.43	b$_b$-minor	6.87	1.19
B-major	11.51	1.47	b-minor	8.35	1.11

partials takes place which results in amplitude modulation (AM) as well as in a certain amount of auditory roughness. A simple method suited to check AM in the complex time signal for each individual chord is measuring the intensity of the sound as a function of time. If AM is present in the signal, the decay curve will show many small fluctuations as are visible in Fig. 5 for the e_b-minor chord in Werckmeister III and the d-minor chord in Vallotti. The modulation frequency and the depth of AM permit a rough assessment of the spectral inharmonicity and the quality of tuning for a certain chord. While the curve of intensity decay is smooth for chords in just tuning (or nearly so), AM increases with deviations from just pitch ratios as well as with spectral inharmonicity corresponding to such deviations.

A signal processing approach suited to investigate AM of individual partials of a complex tone or of a chord comprising several harmonic complexes is the phase vocoder which can be viewed as a filter bank that can be tuned so that the base frequency of the filter bank equals the fundamental of a harmonic complex. For the present study, the sndan software [71, 72] was used which includes tools suited to analyzing AM as well as spectral inharmonicity in harmonic complexes. One of the tools is a 3D-plot of the amplitudes of harmonic partials over time where the temporal envelope for individual partials can be displayed so that AM or other processes become visible. Figure 6 shows partials no. 1–20 from tones in the B-major chord played in Werckmeister III. Figure 7 shows partials 1–20 from the tones in the B-major chord in the Vallotti tuning. B-major is one of the more problematic chords in both tunings (with relatively low HNR readings, see Tables 1 and 2). As is obvious from the graphics displayed, there is considerably AM in both chords. The cause of AM is that, while in just intonation partials from several tones of a major chord played like c_2, g_2, c_3, e_3, g_3 would coincide, in temperaments such as Werckmeister III or Vallotti (or ET12, for that matter) partial frequencies deviate

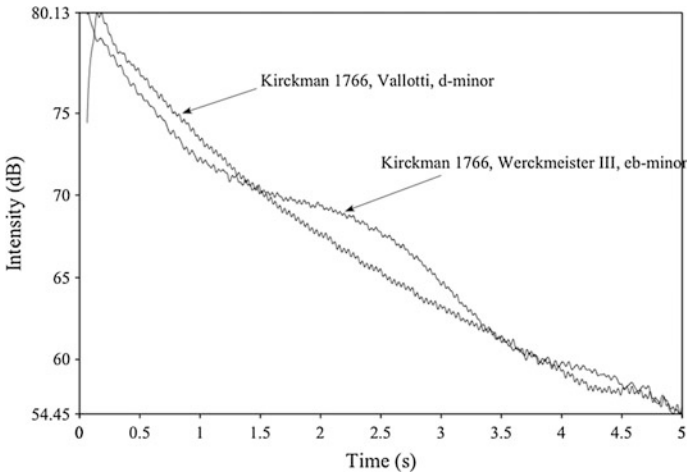


Fig. 5 Decay curves for the e_b-minor chord in Werckmeister III and the d-minor chord in Vallotti show many small fluctuations resulting from harmonic partials undergoing AM

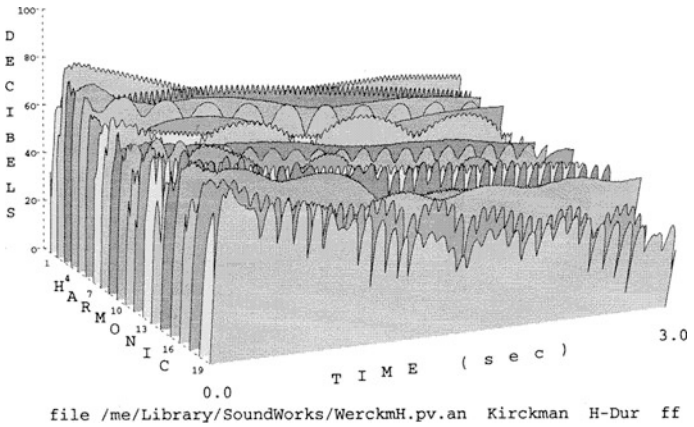


Fig. 6 Werckmeister III, B-major chord, partials 1–20, AM pattern

to some degree from each other (depending on the temperament chosen and the chord that is played). Deviations between pairs or groups of partials can be precisely determined in spectral analysis with appropriate FFT-settings (since $df = fs/N$, where df is the difference limen for two frequency components to be separated, fs is the sampling frequency of the signal, and N is the length of the FFT transform or 'window'). For short FFT windows (e.g., 1024 or 2048 samples per frame), separation is not possible, to the effect that two closely spaced spectral components interact so as to exhibit AM in harmonic plots (see Figs. 6 and 7).

Tools available in sndan furthermore permit to measure the deviation of individual partials from harmonic frequencies as well as to compute such deviation for the weighted average of a number of partials. The results are available as lists

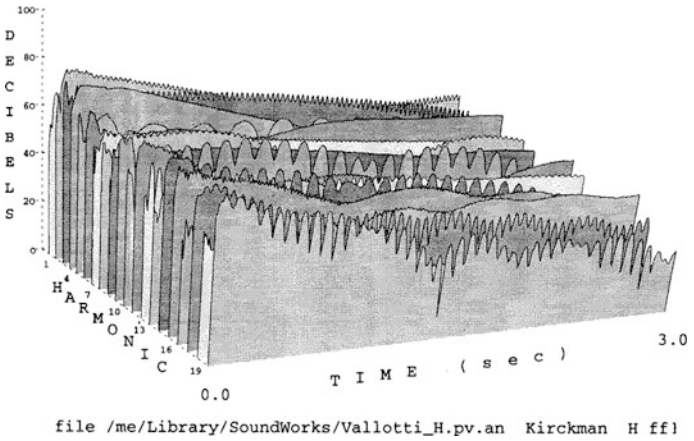


Fig. 7 Vallotti, B-major chord, partials 1–20, AM pattern

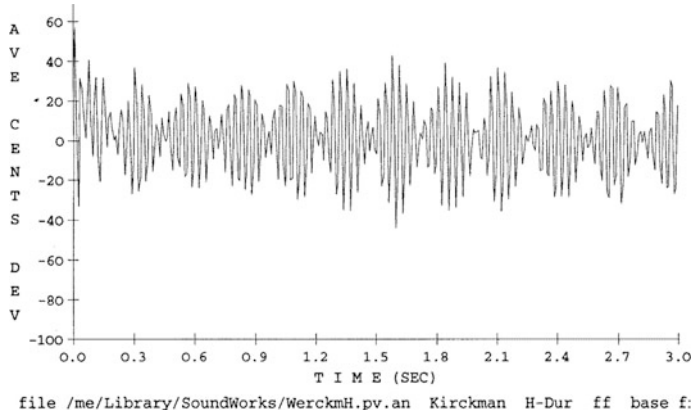


Fig. 8 Werckmeister, B-major chord, weighted average deviation, partials 1–5, cents

(including some statistics) and can be used for a quantitative assessment. Results can also be displayed as a graphic. For example, Fig. 8 shows the weighted average for partials 1–5 of the B-major chord in Werckmeister III and Fig. 9 the same measurement in Vallotti as recorded from the Kirckman.

Deviation at the onset of each sound (0–100 ms) results from the plucking of strings and is found in all tunings. As Figs. 8 and 9 demonstrate, the B-major chord in Werckmeister III shows smaller deviations on average over the first three seconds of recorded sound than the same chord tuned to Vallotti. To be sure, for the given chord structure, computation of the weighted average for the first five harmonic partials already captures four pairs of corresponding partials, one of which relates to the fifth, and another to the major third in the chord, respectively. Hence, deviations

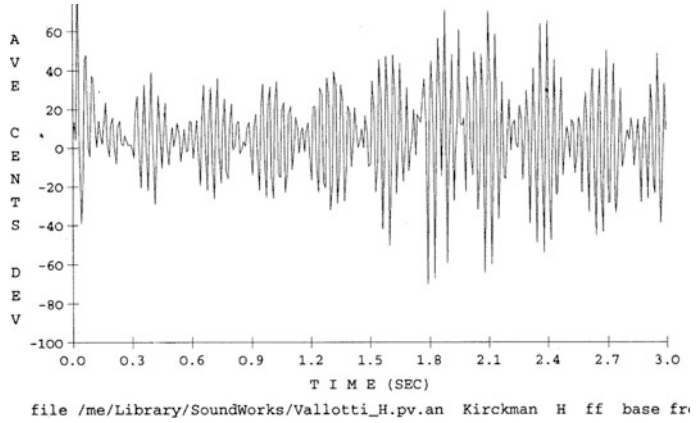


Fig. 9 Vallotti, B-major chord, weighted average deviation, partials 1-, cents

in these two intervals from just ratios account for the computation of the weighted average.

The data from various types of signal analyses may be used for a comparison including statistics. A comparison of the HNR data (Tables 1 and 2) for Werckmeister III and Vallotti might be tempting yet is not easy since the means only indicate the average level (in dB) for certain chords computed from data representing sound segments of a given length. The SD computed for the same data block is considerable for most of the chords, indicating they undergo significant change over time. The HNR in fact goes up with time, for many natural sounds generated by means of an initial impact causing energy transfer into a vibrating system (such as a string that is plucked or a membrane that is struck), because the dissipation of energy due to radiation of sound leads to the rather fast damping of higher partials, meaning the number of partials that can cause inharmonicity (or jitter) in a complex sound such as a chord played on a harpsichord tuned to some temperament diminishes with time. Hence, a certain amount of the variance expressed as SD for each sound of a major and minor chord in our study is attributable to damping out of higher partials due to sound radiation and energy consumption, which lets the HNR rise with time as is shown for three major chords ($F^{\#}$, A_b, A in Werckmeister III) in Fig. 10.

Given these conditions, a weighting of the means of the HNR data by their respective SDs, which can be done by calculating the coefficient of variation (CV) as a statistical parameter, will not be much of help for sounds recorded from plucked strings (while it is a different matter with steady-state sounds recorded from organ pipes). Of course a percentage of the variance in our HNR data results from

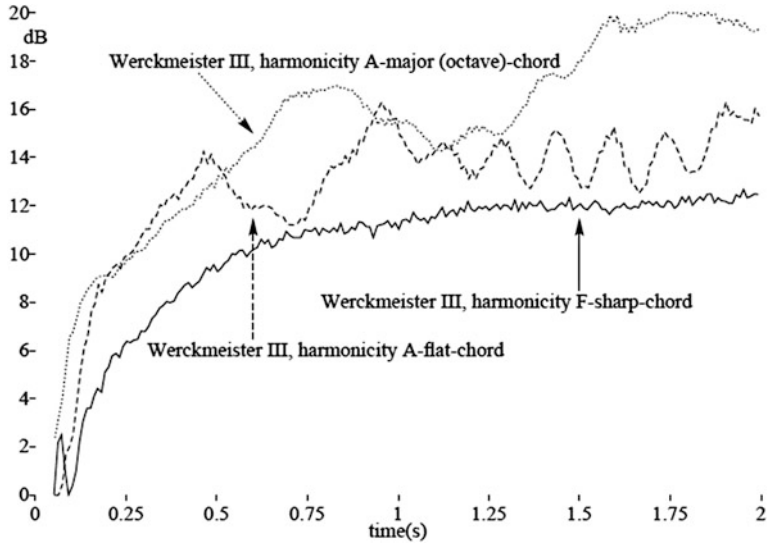


Fig. 10 Evolution of HNR over time for three major chords in Werckmeister III

the relative inharmonicity of partials in chords due to different tunings and, correspondingly, different deviations (in cents) of tones in a given major or minor chord from just intervals. However, the acoustical factor of damping which reduces jitter and "smoothens out" spectral structure in sounds from harpsichord strings also is relevant for explaining the considerably large SD in our HNR data. If one would assume that the damping is more or less the same for chords played in various temperaments (what in fact is not quite true, see for example the decay curves displayed in Fig. 5), one might compare only the means computed for each chord in individual tunings as well as taking the sums of HNR readings for all major and minor chords. In so doing, a small but recognizable advantage of Vallotti (Major chords, $\sum = 160.22$ dB, minor chords $\sum = 93.72$ dB) over Werckmeister III (Major chords $\sum = 152.22$ dB, minor chords $\sum = 91.15$ dB) may be seen. However, comparison in pairs of corresponding major as well as minor chords shows Werckmeister prevails in some of the keys, and Vallotti in others. This is what one would expect from temperaments that, with only 12 tones and pitches to the octave, cannot but seek to install a compromise tuning suited to perform music in all common major and minor keys without producing too much of auditory roughness or even audible mistuning of intervals and chords.

5 Perceptual and Aesthetic Aspects

Keyboard temperaments and tunings have been an issue since the early Renaissance in Europe when the medieval practice of 'Pythagorean' tuning did no longer fit the interval and chord structure developed in a growing number of musical works. Many historical sources on organology, tuning and temperament from the 16th, 17th, and 18th century, respectively, clearly indicate that musicians and also listeners were sensitive to beats and roughness arising from poor tunings. One effect reported quite often was that chords played on keyboard instruments tuned to some temperament did not fit well to melodic lines and polyphonic textures coming from singers, and would interfere in particular with brass instruments most of which were without valves, slides, or keys and thus producing only natural tones of the harmonic series. For example, Mattheson [73, 143–149] who supported equal temperament for keyboard instruments because he saw its advantages (most of all, modulation through all keys), argued that all semitones in ET12 would sound out of tune if compared to the actual intonation of singers, and in particular if compared to the pitches of the trumpet and similar instruments. In an interesting rational discussion of the pros and cons of equal temperament and tuning, he said introducing ET12 in church music would meet grave resistance, first of all, for the sheer number of organs that would need to be retuned (from meantone to ET12), second, in regard of the costs this would generate for each parish. As a third reason, Mattheson pointed to the organ-builders who he said were stubborn and unwilling to let theorists teach them how to tune an organ. A fourth factor according to Mattheson [73, 144] would be the singers and instrumentalists who, after the introduction of

ET12 as a standard tuning, without doubt would sing and play out of tune for some time to come. Finally, as a fifth factor, he mentions the listeners "who didn't yet temper their ears according to the numbers" (such as had been published at the time for ET12). In what looks like an early contribution to the nature vs. nurture debates so common in theories of perception later, Mattheson believed that adaption to equal temperament on the side of listeners would be possible "since habit is the other nature also in this matter".

In several experiments we conducted in the past, samples of subjects were asked to judge harmonic cadences and chord progressions or excerpts from polyphonic pieces of music played in various tunings in regard to perceptual qualities and aesthetic appreciation (see [74–76]). Most of the subjects had musical training, though on different levels of expertise ranging from elementary music education to music academy training as singer, instrumentalist, or conductor. The experiments used various temperaments and tunings (ET12, Vallotti/Young, 1/4-comma meantone, Werckmeister III, Pythagorean, Kirnberger III, as well as selections of just intonation pitches from a 2D tone lattice). Also included in some experiments was the effect of a transposition of a piece from one key into another while the 1/4-comma meantone tuning remained unchanged. Furthermore, the general sensitivity of subjects for defects in tuning was checked by shifting a melodic line 50 cents up or down in pitch while the harmonic accompaniment was left unchanged (cf. [74]). Experimental data subjected to statistical analysis demonstrate that subjects in general are capable of distinguishing temperaments and tunings which they evaluate in regard to perceptual qualities such as consonance, on the one hand, and auditory roughness, on the other. In several of our experiments, subjects were asked to evaluate items also in regard to correctness of musical syntax. Furthermore, in some experiments subjects rated their aesthetic appreciation of musical excerpts played in different tunings. In the following, some of the aforementioned aspects will be addressed with reference to hitherto unpublished data from previous experiments.

In psychoacoustics, it is a common investigation making subjects judge optimal interval sizes. There are several experimental procedures for such tests, for example, one may use two signal generators one of which delivers a signal at a fixed frequency (if the signal is a sine tone) or fundamental frequency (if the signal is a harmonic complex), while the output of the second generator is varied in frequency either by the experimenter or by the subject so that the subject perceives the musical interval (say, a major third) as 'just' or 'perfect'. Typically, musically trained subjects are capable of matching two signals so that their frequencies are in small integer ratios (or nearly so, see, e.g. [77]). The sensory factor most relevant for these judgements is that auditory beats and roughness disappear if two periodic signals are in harmonic frequency ratios for musical intervals such as the fifth, fourth, or major third. Even in successive intervals (tone A followed by tone B) small integer ratios are prevalent as can be tested for the octave; the opinion according to which octaves must be "stretched" to appear as correct in regard to an optimum interval size was not confirmed in a series of experiments we conducted using a standard experimental setup (offered in [78]; see [79, 482–484; 21]).

The general sensitivity of subjects for melodic phrases and/or sequences of chords where some or all tones are out-of-tune can be checked in experiments with variables relating to sensory consonance and dissonance, respectively. In one experiment (2006), 50 students of the University of Hamburg were asked to rate the consonance and dissonance they perceived with musical stimuli on scales ranging from 1 (low) to 7 (high). The variables in this experiment were (1) consonance, (2) dissonance, (3) goodness of intonation (German: Intonationsgüte), and (4) aesthetic overall impression. Measures 1–8 from Bach's Invention no. 1 (C-Major, BWV 772) served as a musical stimulus, from which several variants were produced with different tunings and sounds. Version 1 has the sound of a harpsichord tuned to ET12. The sound comes from FM synthesis (Yamaha TX 81 Z) and appears realistic in regard to temporal and spectral features. Version 2 has the same sound but employs Vallotti/Young tuning. Version 3 again is in ET12 but has a special sound synthesized from components spaced in octaves and played from a hardware sampler (SE synthesis on an EMAX II stereo). Version 4 has the harpsichord sound and ET12 tuning yet with a stretch of 50 cent between notes of the voices in the two-part invention. Hence all simultaneous intervals are too wide by a margin of 50 cents (a quarter of a whole tone in ET12). Version 5 was based on a selection of 12 pitches from a 2D lattice (tone net) comprising fifths and major thirds in just intonation, played with the harpsichord sound used also in versions 1, 2, and 4. Subjects were asked to rate both consonance and dissonance as two variables, which not only generates additional data but permits a more precise assessment of the perceptions subjects have from the stimuli. Of course, these two variables interrelate closely (meaning high ratings for consonance should go along with low ratings for dissonance, and vice versa). The descriptive statistics for the five versions are listed in Table 3.

Without going into a detailed analysis of the data at this place, one can see that version 2 in Vallotti/Young received best ratings on three of the four variables, and that version 4 was perceived as clearly out-of-tune by the subjects in the sample as is evident from low ratings for consonance and for goodness of intonation as well as for overall aesthetic impression while ratings for dissonance are much higher in this version than in any other. We may conclude from these figures that detuning tones in simultaneous intervals in a musical setting by as much as 50 cents will have strong perceptual and aesthetic effects on listeners. However, ratings for version 1 in ET12 and for version 2 (Vallotti/Young) differ not significantly for the variables based on sensory qualities (consonance, dissonance) while the difference for goodness of intonation is more marked, and that for aesthetic overall impression

Table 3 Means, SDs for Invention no. 1, 5 versions (2006, n = 50 subjects)

Variable	Version 1	Version 2	Version 3	Version 4	Version 5
Consonance	5.73, 1.22	5.81, 1.1	5.46, 1.22	2.56, 1.75	5.0, 1.4
Dissonance	1.96, 1.18	1.91, 1.06	2.29, 1.22	5.35, 1.9	2.83, 1.46
Intonation goodness	5.20, 1.59	5.43, 1.37	4.78, 1.57	2.24, 1.8	4.24, 1.39
Aesthetic impression	2.74, 1.41	3.04, 1.38	3.36, 1.41	1.44, 0.73	2.60, 1.24

still somewhat greater as is indicated by median values (ET12 = 2, Vallotti/Young = 3; this difference though is not large enough to yield significant results in an U-test where z = 1.146 (1.960, p = 0.05). From the results of this experiment one may conclude that musical excerpts played in either ET12 or in Vallotti/Young differ not significantly in regard to perceptual effects and aesthetic appreciation even though Vallotti/Young, in a direct comparison such as performed in this experiment, prevails. The somewhat higher ratings for Vallotti/Young as are reflected in the judgements could be attributed to the slightly better figures this tuning achieves in an overall assessment of deviations from just intonation (cf. [63–65]).

Another comparative evaluation can be made from data of an experiment (2001) in which a sample of 44 subjects (all students in their first or second semester) listened to measures 1–15 from J.S. Bach's Sonata in E_b-Major (BWV 552) in five different tunings, namely (1) Vallotti/Young, (2) 1/4-comma meantone tuned from c as base note, (3) ET12, (4) 1/4-comma meantone with the scale based on e_b, (5) Kirnberger III. The five musical excerpts were performed with a synthesized pipe organ sound (TX 81 Z) and were judged on the dimensions (a) consonance, (b) roughness, and (c) goodness of intonation (German: Intonationsgüte). The design can be stated as one factor (tunings) with 5 conditions. The descriptive statistics for the data are summarized in Table 4.

Table 4 Judgement of different tunings, n = 44 subjects, Hamburg 2001

Tuning	Data file no.	Mean	SD	Median	Range	CV (%)
(1) Vallotti/Young						
Consonance	1	4.55	1.21	4.5	2–7	26.6
Roughness	2	3.36	1.51	3	1–7	44.92
Intonation	3	3.98	1.23	4	2–7	30.91
(2) Meantone (c)						
Consonance	4	4.05	1.38	4	2–6	34.12
Roughness	5	3.77	1.63	3.5	1–7	43.11
Intonation	6	3.16	1.45	3	1–7	45.77
(3) ET12						
Consonance	7	4.96	1.14	5	2–7	23.02
Roughness	8	3.32	1.44	3	1–6	43.49
Intonation	9	4.34	1.35	4	1–7	31.0
(4) Meantone (e_b)						
Consonance	10	4.96	1.43	5	2–7	28.86
Roughness	11	3.55	1.56	3	1–7	44.06
Intonation	12	4.21	1.61	4	1–7	38.24
(5) Kirnberger III						
Consonance	13	4.8	1.46	5	1–7	30.36
Roughness	14	3.8	1.62	3	1–7	42.74
Intonation	15	4.1	1.76	4	1–7	42.84

Data files were checked in groups representing the three variables for homogeneity of variances with Bartlett-tests. Since variances are sufficiently homogeneous, an ANOVA conducted for consonance (data files 1, 4, 7, 10, 13) yields $F = 3.628$ ($F[0.01] = 3.418$), which is very significant, but for roughness (data files 2, 5, 8, 11, 14) ANOVA is not significant at $F = 0.901$. ANOVA for goodness of intonation (data files 3, 6, 9, 12, 15) yields $F = 4.301$, which again is very significant. Following the ANOVA, a multiple-mean test (Scheffé) was conducted for each variable, which yields significant contrasts between data files for consonance (D 4/7 and D 4/10, $p < 0.025$), and between data files for the goodness of intonation (D 6/12, $p < 0.05$, D 6/9, $p < 0.01$). Hence, the meantone scales tuned from either c or e_b make a perceptual difference for a piece of music in the key of E_b. For the data files representing five tunings, also a MANOVA can be computed, taking consonance, roughness and goodness of intonation as three variables dependent on the factor (tunings). MANOVA yields Wilks-$\lambda = 0.57$, $F = 8.11$ ($F[0.001] = 3.113$), which underpins the tunings differ in regard to perceptual and musical effects. However, an inspection of the descriptive statistics shows that the SD and the range for all variables is large (as is the CV for variables in several tunings), indicating that subjects in the sample varied markedly in their individual judgements. There are several possible explanations for these figures, one of which is that the subjects in the sample were young students not experienced with different tunings and apparently quite uncertain in their judgments. Furthermore, the sample is not homogeneous and in fact contains several groups of subjects that differ in regard to their musical ability, education, and preferences. This holds true even for students in musicology where, besides individuals with a conventional 'classical' music training, today one finds a growing number of young people who come from jazz, rock/pop, electronic music genres or (depending on their family history as migrants) various non-western music cultures. In particular students with a musical background predominantly in rock/pop/electronic genres are used to ET12 as this is the standard tuning not only on keyboards but also on fretted string instruments (guitars, bass). In addition, these subjects are used to sounds that are heavily processed with effect units, many of which involve temporal and spectral modulation (see [80–82]). For example, spectral envelope and energy distribution can be modulated with a bandpass filter where the center frequency and the bandwidth vary with time. In a phaser circuit, a low-frequency oscillator (LFO) controls the amount of time by which a signal is delayed relative to the dry input signal. Adding the delayed signal to the dry signal, constructive and destructive interference results depending on the phase angle of the delayed signal relative to the dry one. Since the delay is varied with time, the phase angle between the two signals also varies periodically as does the spectral energy distribution, giving sounds processed in this way a "breathy" timbral quality. Some effects such as chorus and flanger can be set so that the pitch of a sound is modulated periodically up and down the fundamental frequency of the input signal. Effect units which modulate spectral and temporal characteristics are employed not the least to give synthesized or sampled sounds played on a keyboard instrument in ET12 a lively quality. Also, effects such as chorus and delay effects are used to double and 'broaden' pitches for singers who

thus seem to possess of a 'big voice'. Singing with a chorus effect (instead of precise intonation) has become 'industrial standard' in many pop productions. Apparently, young students already in the 1990s were so used to sounds undergoing permanent modulation that, in some of our experiments, they rated harmonic major cadences or musical pieces played with a synthesized pipe organ or similar harmonic complex sound and tuned to ET12 higher with respect to 'pleasantness' than any tuning that produced a high degree of sensory consonance yet seemed static in regard to both pitch and spectrum. For instance, among the music examples we employed in several experiments was a polyphonic setting of the chorale *Wie schön' leucht uns der Morgenstern* (BWV 763), of which Reinier Plomp [31] provides an excerpt (measures 1–6) in two versions, one in just intonation, and another in ET12. The version in just intonation offers a high degree of consonance (for the coincidence of many partials and the lack of beats and roughness) but is a bit sharp in timbre (again, for the coincidence of many partials which brings the spectral centroid up to higher frequencies). In contrast, the version in ET12 is less transparent yet may appear "warm" in sound quality because of the interference of partials as well as the various modulation products which become audible over time. The response of young students (with a major in systematic musicology or in other subjects such as sociology, media science, etc.), and in particular of those with a background in pop/rock/electronic music typically was that they rather preferred the ET12 version as this apparently was close to the sound quality they had experienced in music genres of their choice. Given these changes in listening attitudes and preferences, Mattheson (1731, see above) perhaps was right when he argued that "habit is the other nature".

6 Conclusion

A review of historical sources (e.g., [4, 8, 12]) shows tuning and temperaments was a major issue in music theory and organology in the time from, roughly, 1400 to 1900. Besides more theoretical elaborations, there are many sources which clearly indicate musicians and instrument builders experimented with different tunings to find solutions for a discrepancy that stems from the nature of musical intervals being governed by different prime numbers (2, 3, 5). In addition to the octave, the perfect fifth and the fourth that had ranked as 'symphonous' in medieval music theory (see [10]), the just major third and other harmonic intervals were included into composition and performance from c. 1450–1500 on, forcing music theorists and instrument builders to deal with this new situation. Accepting the just major third as a fundamental interval bears implications both in regard of the division of the fifth into major and minor third as well as the division of the major third into greater and lesser whole tone. Extending the range of usable harmonic intervals to the just major and minor third and the just major and minor sixth means appropriate tones or 'pitches' in a scale must be available if the same intervals shall be played, on a keyboard instrument, in various keys. With only 12 tones and pitches to the

octave, one can tune these so that *some* major and minor chords are in just intonation, that is, their fifths are perfect and their major and/or minor thirds are in just frequency ratios (the equivalent to string sections on a monochord, as shown by Ramis de Pareia in [40]). Just intonation plus capability to modulate among several keys (both sharps and flats) inevitably leads to more than 12 tones and pitches to the octave. If one wants to build a keyboard instrument in just intonation, the number of tones and pitches required per octave and the actual selection of pitches and tones in a tuning depends on the range of keys that shall be covered and the degree of justness that is deemed appropriate (see [35]).

A temperament in certain respects is a means to reduce the larger number of pitches that would be needed, in just intonation (perfect fifths and fourths, just thirds and sixths), to a number considerably smaller but still sufficient to realize pitches so that they are relatively close to the just intonation pitches they 'represent'. Let m be the large number of tones and pitches in a 2D-lattice of just intonation, and n the number of pitches and tones available from a certain temperament, where $m \gg n$. Since technology and playability impose restrictions on the design of conventional keyboard instruments, the problem here is to find the smallest number n suited to 'represent' as many of the tones and pitches m as are deemed necessary by a composer (taking the notation in individual works of music as a source for analysis). For example, the Duetto I (BWV 802) from the 3rd part of Bach's *Clavier-Übung*, even though it is only a two-part musical setting, has no less than 17 different notes in the score (supervised for print by Bach himself), which express musically distinct intervals as intended by the composer. If the Duetto I were to be played in a tuning suited to keep different intervals distinct as simultaneous sonorities, more than twelve tones and pitches to the octave will be needed. In other advanced organ and harpsichord works of Bach like the Fantasie und Fuge in g-minor (BWV 542) or the Chromatische Fantasie und Fuge in d-minor (BWV 903), the number of pitches and tones found in the notation as well as by an analysis of the harmonic structure is considerably higher than 17.

There have been various attempts at finding an optimal relation for $\{m, n\}$, one of which is a division of the octave into equal parts, where n can be any whole number such as 72 (see [55]) or 31 (the Cycle Harmonique of Huygens from 1661/1691, see [6, 187ff.]). If the aim is to realize all thirds (5/4) and fifths (3/2) with very good approximations, $n = 53$ will be chosen (as calculated first by Mercator and Holder and advocated also by Helmholtz and Bosanquet, in the 19th century). Though some keyboard instruments have been manufactured with 53 pitches and keys to the octave, mainly for experimental purposes, it is not a very practical solution for a harpsichord maker (and less so for an organ builder). In case one wants just thirds and sixths and is willing to accept the somewhat narrowed fifths and widened fourths of Huygens' cycle, then $n = 31$ will do. This is a solution that has been implemented (either as an equal temperament, or with some variation in interval size, see [3, 7, 33]) on several keyboard instruments from the 16th century to our times (the 31-tone pipe organ conceived by Fokker was installed in 1950, and in the early 1970s, an electronic keyboard with the Huygens-Fokker-tuning was built for Webster College in St. Louis). A division of the octave into 19 equal parts has also been discussed at

times, but the benefit from $n = 19$ in regard to approximating just intonation pitches will be quite small. Finally, confining the number of tones and pitches per octave to $n = 12$, several models for tuning a temperament are feasible (cf. [11, 17]), one of which is ET12, another is Werckmeister III, and still another is the scale and tuning model devised by Vallotti and by Young. The 1/4-comma meantone model, which in general is regarded as a temperament, in fact is a mixture of just intonation intervals (taking the eight just major thirds and seven just minor thirds that can be realized in this tuning, see above) and a tempering of the fifths and fourths as well as a division of the major third into two meantones of equal size.

The pros and cons of various temperaments have been discussed extensively in works on tuning and temperament, often relying on personal experience of musicians and theorists as well as on reports from organ builders or music experts that were called to examine new organs (like J.S. Bach). Though such reports are valuable as historical sources, an objective assessment of tunings and temperaments by means of computing deviations from just intonation intervals allows quantifying the goodness-of-fit of various temperaments (cf. [63, 65]). In addition, examination of tunings on the basis of actual sound recordings of real or synthesized instruments subjected to signal analysis seems necessary since the quality of a tuning for musicians and listeners depends on factors such as periodicity in the time domain and spectral harmonicity of partials in the frequency domain (cf. [18, 28, 29, 67]). In the present study, these parameters have been investigated, to some extent, for the Werckmeister III and the Vallotti temperament tuned on a historical Kirckman harpsichord by computing the HNR for signals recorded for major and minor chords played in the aforenamed temperaments. Furthermore, we reported empirical data from some experiments where subjects were asked to rate musical excerpts played in several tunings such as 1/4-comma meantone, Werckmeister III, Vallotti/Young, and ET12. The data suggest that subjects with some musical training can distinguish between different tuning models in case their differences are large enough to have effects for perception that can be measured on psychoacoustic variables (consonance, dissonance, roughness). The effects are less marked, though, for aesthetic appreciation where in particular young students nowadays seem to prefer ET12 because of their intensive exposure to sounds from electronic keyboards including digital audio effects employed for spectral and pitch modulation.

Acknowledgment We like to thank James Beauchamp for continuous access to his sndan package.

References

1. Ortgies, I.: Die Praxis der Orgelstimmung in Norddeutschland im 17. und 18. Jahrhundert und ihr Verhältnis zur zeitgenössischen Musikpraxis. Ph.D. Dissertation, University of Göteborg (2004)
2. Wraight, D.: The *cimbalo cromatico* and other Italian string keyboard instruments with divided accidentals. Schweizer Jahrb. Musikwiss. **22**, 105–136 (2003)

3. Barbieri, P.: Enharmonic Instruments and music 1470–1900. Levante, Latina (2008)
4. Barbour, J.M.: Tuning and temperament. East Lansing, MI, Michigan State University Press (repr. New York: DaCapo Pr. 1972) (1951/1972)
5. Barker, A.: Greek Musical Writings, vol 2: Harmonic and Acoustic Theory. Cambridge University Press, Cambridge (1989)
6. Fokker, A.D.: Rekenkundige Bespiegeling der muziek. Noorduijn, Gorinchem (1945)
7. Fokker, A.D.: Neue Musik mit 31 Tönen. Düsseldorf: Verlag der Ges. zur Förderung der Syst. Musikwiss. (engl. transl. New Music with 31 notes. Bonn: Verlag für Syst. Musikwiss. 1975) (1966/1975)
8. Lindley, M.: Stimmung und Temperatur. In: Dahlhaus, C. et al. (eds.) Hören, Messen und Rechnen in der frühen Neuzeit (=Geschichte der Musiktheorie, Bd 6). Wiss. Buchges., Darmstadt, pp. 109–331 (1987)
9. Lindley, M., Turner-Smith, R.: Mathematical Models of Musical Scales. Verlag für Syst. Musikwiss, Bonn (1993)
10. Münxelhaus, B.: Pythagoras musicus. Zur Rezeption der pythagoreischen Musiktheorie als quadrivialer Wissenschaft im lateinischen Mittelalter. Verlag für Syst. Musikwiss, Bonn (1976)
11. Rasch, R.: Description of regular twelve-tone musical tunings. J. Acoust. Soc. Am. **73**, 1023–1035 (1983a)
12. Ratte, F.J.: Die Temperatur der Clavierinstrumente. Bärenreiter, Kassel (1991)
13. Sachs, K.J.: Mensura fistularum. Die Mensurierung der Orgelpfeifen im Mittelalter, Bd 2. Musikwiss. Verlagsges, Murrhardt (1980)
14. Vogel, M.: Die Lehre von den Tonbeziehungen. Bonn: Verlag für Syst. Musikwiss. (engl. transl. On the Relations of tone. Bonn: Verlag für Syst. Musikwiss. 1993) (1975/1993)
15. Barnes, K.: Bach's keyboard temperament. Internal evidence from the *Well-tempered Clavier*. Early Music **7**, 23–249 (1979)
16. Billeter, B.: Bachs Klavier- und Orgelmusik. Amadeus, Winterthur (2010)
17. Blackwood, E.: The Structure of Recognizable Diatonic Tunings. Princeton University Press, Princeton (1985)
18. Norrback, J.: A Passable and Good Temperament. A New Methodology for Studying Tuning and Temperament in Organ Music. Institute of Musicology, Göteborg (2002)
19. Werckmeister, A.: Musicalische Temperatur. Calvisii, Quedlinburg (repr. Utrecht: Diapason Pr. 1983) (1691/1983)
20. Kirnberger, J.P.: Die Kunst des reinen Satzes. T. I and II. Lagarde, Berlin (1771)
21. Schneider, A.: Psychophysics and psychoacoustics. In: Bader, R. (ed.) Handbook of Systematic Musicology. Springer, Berlin, New York (2016)
22. Wiener, N.: Generalized harmonic analysis. Acta Math. **55**, 117–258 (1930)
23. Khintchine, A.: Korrelationstheorie der stationären stochastischen Prozesse. Math. Ann. **109**, 604–615 (1934)
24. Hartmann, W.: Signals, Sound, and Sensation. AIP Press and Springer, New York (1998)
25. Meyer, E., Guicking, D.: Schwingungslehre, 2nd edn. Vieweg, Braunschweig (1974)
26. Beurmann, A., Schneider, A.: Acoustics of the harpsichord: a case study. In: Schneider, A. (ed.) Systematic and Comparative Musicology: Concepts, Methods, Findings, pp. 241–263. Peter Lang, Frankfurt am Main (2008)
27. Beurmann, A., Schneider, A.: Acoustics and sound of the harpsichord: another case study. In: Bader, R. (ed.) Musical Acoustics, Neurocognition and Psychology of Music, pp. 57–72. Peter Lang, Frankfurt am Main (2009)
28. Schneider, A., von Ruschkowski, A., Bader, R.: Klangliche Rauhigkeit, ihre Wahrnehmung und Messung. In: Bader, R. (ed.) Musical Acoustics, Neurocognition and Psychology of Music, pp. 103–148. Peter Lang, Frankfurt am Main (2009)
29. Sethares, W.: Tuning, Timbre, Spectrum, Scale, 2nd edn. Springer, London (2004)
30. Plomp, R.: Aspects of Tone Sensation. Academic Press, London (1976)
31. Plomp, R.: Hoe wij horen. Over de toon die de muziek maakt [How We Hear. Concerning the Tone that Makes Music. CD + two booklets]. Author, Breukelen, NL (1998)

32. Benson, D.: Music. A Mathematical Offering. Cambridge University Press, Cambridge (2006)
33. Cordes. M.: Nicola Vicentinos Enharmonik. Musik mit 31 Tönen. Akademische Druck- und Verlagsanstalt, Graz (2007)
34. Tanaka, Sh: Studien im Gebiete der reinen Stimmung. Vierteljahrsschr. Musikwiss. **6**, 1–90 (1890)
35. Groven, E.: Temperierte und Reine Stimmung. Author, Oslo, Suttung (1973)
36. Rudi, J.: Eivind Groven and technological innovation. In: Loe Dalaker, I., Jorunn Kydland, A., Łopatowska-Romsvik, D. (eds.) 'East of Noise'. Eivind Groven. Composer, Ethnomusicologist, Researcher. Akademika Publications, Oslo, Trondheim, pp. 173–190 (2013)
37. Düring, I.: Die Harmonielehre des Klaudios Ptolemaios (Göteborgs Högskolas Årsskrift T. 36, no. 1). Wettergren & Kerbers, Göteborg (1930)
38. Huffman, C.: Archytas of Tarentum. Pythagorean Philosopher and Mathematician King. Cambridge University Press, Cambridge (2005)
39. Klotz, H.: Über die Orgelkunst der Gotik, der Renaissance und des Barock, 2nd edn. Bärenreiter, Kassel (1975)
40. Ramis de Pareia, B.: Practica musica. Bologna (ed. J. Wolf, Leipzig: Breitkopf Haertel 1901) (1482/1901)
41. Schlick, A.: Spiegel der Orgelmacher und Organisten. Author, Speyer (repr. Mainz: Schott 1974) (1511/1974)
42. Dupont, W.: Geschichte der musikalischen Temperatur. Bärenreiter, Kassel (1935)
43. Adamczyk, S.: Das Stimmungssystem des Johannes von Lublin (1540). Ars Organi. **51**, 224–227 (2003)
44. Glarean, H.: Glareani ΔΩΔΕΚΑΧΟΡΔΟΝ [Dodekachordon]. Basileae: (Faks. ed. Hildesheim: Olms 1969) (1547)
45. Powers, H.: From psalmody to tonality. In: Judd, C.C. (ed.) Tonal Structures in Early Music, pp. 275–340. Garland, New York (1998)
46. Aaron, P.: Toscanello in musica, 3rd edn. Venezia (repr. Kassel: Bärenreiter 1970) (1523/1539)
47. Rasch, R.: Why were enharmonic keyboards built? From Nicola Vicentino (1555) to Michael Bulyowski (1699). Schweizer Jahrb. für Musikwiss. **22**, 35–93 (2003)
48. Zarlino, G.: Le Istitutioni harmoniche. No publ, Venezia (3rd edn. 1573; repr. of the 1st ed. 1558 New York: Broude 1965) (1558/1573)
49. Praetorius, M.: Syntagma musicum, T. II: De organographia. Holwein, Wolffenbüttel (repr. Kassel: Bärenreiter 1966) (1619/1966)
50. Vicentino, N.: L'antica musica ridotta alla moderna prattica. A. Barre, Roma (1555)
51. Zarlino, G.: Dimostrazioni harmoniche. Francheschi, Venezia (1571)
52. Zarlino, G.: Sopplimenti musicali. Francheschi, Venezia (1588)
53. Fogliano, L.: Musica theorica docte simul ac dilucide pertracta. G.A. Nicolini, Venezia (repr. New York: Broude 1969) (1529)
54. Zarlino, G.: Modulationes sex vocum. Venezia. Critical ed. by C. Collins Judd and K. Schiltz. A-R Editions, Madison, WI, 2015 (1566/2015)
55. Maedel, R., Richter Herf, F.: Ekmelische Musik. Katzbichler, München, Salzburg (1977)
56. Rameau, J.P.: Nouveau système de musique théorique. Ballard, Paris (1726)
57. Rameau, J.P.: Génération harmonique ou traité de musique. Prault fils, Paris (1737)
58. Heinichen, J.: Der General-Bass in der Composition. No publ, Dresden (1728)
59. Heinichen, D.: Neu erfundene und gründliche Anweisung, wie ein Music-liebender auff gewisse vortheilhaftige Arth könne zu vollkommener Erlernung des General-Basses … gelangen. Schiller, Hamburg (1711)
60. Rasch, R.: Introduction to A. Werckmeister. Musicalische Temperatur (Quedlinburg 1691). Repr. Utrecht: Diapason Pr., pp. 8–51 (1983b)
61. Lehman, B.: Bach's extraordinary temperament: our Rosetta stone (part 1 and 2). Early Music **33**, 3–23 and 211–231 (2005)
62. Wood, A.: The Physics of Music, 5th edn. (rev. by J.M. Bowsher). Methuen, London (1962)

63. Hall, D.: The objective measurement of goodness-of-fit for tunings and temperaments. J. Music Theor. **17**, 274–290 (1973)
64. Hall, D.: Musical Acoustics. An Introduction. Wadworth, Belmont, CA (1980)
65. Hall, D.: Quantitative evaluation of musical scale tunings. Am. J. Phys. **42**, 543–552 (1974)
66. Boersma, P.: Accurate short-term analysis of the fundamental frequency and the harmonic-to-noise ratio of a sampled sound. In: Proceedings of Institute of Phonetic Science, University of Amsterdam, vol. 17, pp. 97–110 (1993)
67. Schneider, A., von Busch, R.: Zur Verwendung der Viertelkomma-Mittelton-Stimmung auf historischen Orgeln: einige empirische Daten und musikalisch-akustische Kriterien. Acta Organologica **34**, 437–454 (2015)
68. Beurmann, A.: Harpsichords and More. Harpsichords, Spinets, Clavichords, Virginals. Olms, Hildesheim, New York (2012)
69. Beurmann, A., Bader, R., Schneider, A.: An acoustical study of a Kirkman harpsichord from 1766. Galpin Soc. J. **63**, 61–72 (2010)
70. von Helmholtz, H.: Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik, 3rd edn. Vieweg, Braunschweig (1870)
71. Beauchamp, J.: New methods for computer analysis and synthesis of musical sounds. In: Proceedings of 32nd Czech Conference on Acoustics, Prague, 7–15 Sept 1995
72. Beauchamp, J.: Analysis and synthesis of musical instrument sounds. In: Beauchamp, J. (ed.) Analysis, Synthesis, and Perception of Musical Sounds, pp. 1–89. Springer, New York (2007)
73. Mattheson, J.: Grosse General-Bass-Schule …, 2nd edn. Kissner, Hamburg (1731)
74. Schneider, A.: Tone system, intonation, aesthetic experience: theoretical norms and empirical findings. Systematische Musikwiss. (Syst. Musicology) **2**, 221–254 (1994)
75. Schneider, A.: On categorical perception of pitch and the recognition of intonation variants. In: Pylkkänen, P., Pylkkö, P., Hautamäki, A. (eds.) Brain, Mind, and Physics, pp. 250–261. IOS Press & Ohmsha, Amsterdam, Oxford, Tokyo, Washington, DC (1997)
76. Schneider, A.: Über Stimmung und Intonation. Systematische Musikwiss. (Syst. Musicology) **6**, 27–49 (1998)
77. Reinecke, H.P.: Experimentelle Beiträge zur Psychologie des musikalischen Hörens. Sikorski, Hamburg (1964)
78. Houtsma, A., Rossing, Th, Wagenaar, W.: Auditory Demonstrations. Philips & Acoust. Soc. Am, Eindhoven, New York (1987)
79. Schneider, A.: Tonhöhe – Skala – Klang. Akustische, tonometrische und psychoakustische Studien auf vergleichender Grundlage. Orpheus, Bonn (1997)
80. Dutilleux, P., Zölzer, U.: Filters. In: Zölzer, U. (ed.) DAFX. Digital Audio Effects, pp. 31–62. Wiley, Chichester, New York (2002)
81. Dutilleux, P., Zölzer, U.: Delays. In: Zölzer, U. (ed.) DAFX. Digital Audio Effects, pp. 63–74. Wiley, Chichester, New York (2002)
82. Dutilleux, P., Zölzer, U.: Modulators and demodulators. In: Zölzer, U. (ed.) DAFX. Digital Audio Effects, pp. 75–92. Wiley, Chichester, New York (2002)

Author Biographies

Albrecht Schneider (b. 1949) worked as a professor of systematic musicology in the Institute of Musicology, University of Hamburg, from 1981 to 2011 (and as an emeritus since). He has taught as visiting professor in the Department of Ethnomusicology and Systematic Musicology at UCLA and has lectured also in various European institutions. He is the author, co-author, editor and co-editor of a number of books and has contributed chapters and articles to scientific handbooks, encyclopedias, and journals (for bio-bibliographical data, see R. Bader, Chr. Neuhaus, and U. Morgentern [eds]. *Concepts, Experiments, and Fieldwork: Studies in Systematic Musicology and*

Ethnomusicology. Frankfurt/M., Berne, Brussels: P. Lang 2010, and www.fbkultur.uni-hamburg.de/sm/personen.html).

Andreas Beurmann (1928–2016) studied musicology and physics at the universities of Cologne and Göttingen. After receiving his Ph.D. (with a dissertation on the keyboard works of C.Ph.E. Bach) from Göttingen, in 1953, Beurmann worked as a recording engineer and music producer for the public radio, then founding (together with a former co-student and an American business partner) his own record company, in which he became active both as a producer and as a managing director. He was also involved in the development of electronic studio equipment (including an early hardware sequencer build in cooperation with Wolfgang Palm, the founder of PPG). Beurmann had training as a pianist, from where he expanded to the harpsichord (recording several CDs, the last of which was issued in 2015). His collection of historical and contemporary keyboards comprises more than 200 instruments. Andreas Beurmann, who was awarded a honorary professorship from the University of Hamburg for his engagement in teaching musicology, died in April, 2016.