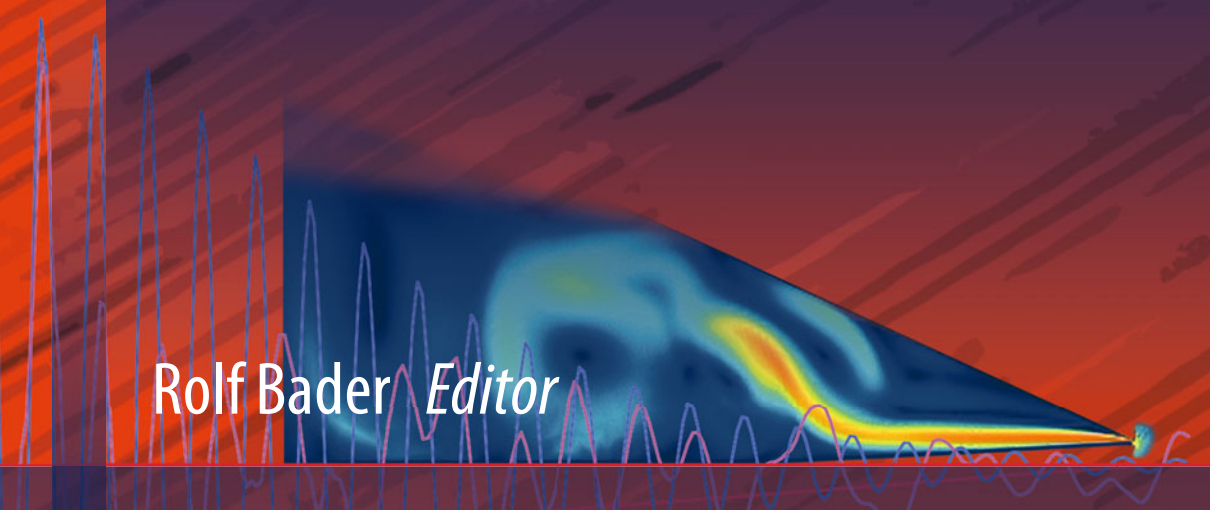
Current Research in Systematic Musicology



Rolf Bader *Editor*

Computational Phonogram Archiving

Springer

Current Research in Systematic Musicology

Volume 5

The series covers recent research, hot topics, and trends in Systematic Musicology. Following the highly interdisciplinary nature of the field, the publications connect different views upon musical topics and problems with the field's multiple methodology, theoretical background, and models. It fuses experimental findings, computational models, psychological and neurocognitive research, and ethnic and urban field work into an understanding of music and its features. It also supports a pro-active view on the field, suggesting hard- and software solutions, new musical instruments and instrument controls, content systems, or patents in the field of music. Its aim is to proceed in the over 100 years international and interdisciplinary tradition of Systematic Musicology by presenting current research and new ideas next to review papers and conceptual outlooks. It is open for thematic volumes, monographs, and conference proceedings. The series therefore covers the core of Systematic Musicology,—Musical Acoustics, which covers the whole range of instrument building and improvement, Musical Signal Processing and Music Information Retrieval, models of acoustical systems, Sound and Studio Production, Room Acoustics, Soundscapes and Sound Design, Music Production software, and all aspects of music tone production. It also covers applications like the design of synthesizers, tone, rhythm, or timbre models based on sound, gaming, or streaming and distribution of music via global networks.

- Music Psychology, both in its psychoacoustic and neurocognitive as well as in its performance and action sense, which also includes musical gesture research, models and findings in music therapy, forensic music psychology as used in legal cases, neurocognitive modeling and experimental investigations of the auditory pathway, or synaesthetic and multimodal perception. It also covers ideas and basic concepts of perception and music psychology and global models of music and action.
- Music Ethnology in terms of Comparative Musicology, as the search for universals in music by comparing the music of ethnic groups and social structures, including endemic music all over the world, popular music as distributed via global media, art music of ethnic groups, or ethnographic findings in modern urban spaces. Furthermore, the series covers all neighbouring topics of Systematic Musicology.

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

Rolf Bader
Editor

Computational Phonogram Archiving

Springer

Editor
Rolf Bader
Institute of Systematic Musicology
University of Hamburg
Hamburg, Germany

ISSN 2196-6966 ISSN 2196-6974 (electronic)
Current Research in Systematic Musicology
ISBN 978-3-030-02694-3 ISBN 978-3-030-02695-0 (eBook)
https://doi.org/10.1007/978-3-030-02695-0

Library of Congress Control Number: 2018958359

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

Contents

Part I Overview

Computational Music Archiving as Physical Culture Theory 3
Rolf Bader

Part II Fieldworks and Archives

"The *Lanang* Is the Bus Driver": Intersections of Ethnography and Music Analysis in a Study of Balinese *Arja* Drumming 37
Leslie Tilley

Temperament in Tuning Systems of Southeast Asia and Ancient India 75
Rolf Bader

John Blacking Revisited—Comparative Analysis of Venda Tshikone Dance (1958 and 2009) 109
Jukka Louhivuori

Smithsonian Folkways and the Associated Ralph Rinzler Folklife Archives and Collections 121
Jeff Place

Analysis and Perception of Javanese *Gamelan* Tunings 129
Gerrit Wendt and Rolf Bader

Part III Computation, Networking and Platforms

Content-Based Music Retrieval and Visualization System for Ethnomusicological Music Archives 145
Michael Blaß and Rolf Bader

Spatial Manipulation of Musical Sound: Informed Source Separation and Respatialization 175
Sylvain Marchand

Finding Music in Music Data: A Summary of the DaCaRyH Project 191
Oded Ben-Tal, Bob L. Sturm, Elio Quinton, Josephine Simonnot and Aurelie Helmlinger

Experimental Investigations and Future Possibilities in Network-Mediated Folk Music Performance 207
Chrisoula Alexandraki

Requirements and Use Cases for Digital Sound Archives in Ethnomusicology 229
Jonas Franke

Part IV Physical Modeling and Measurements

Laser-Based Interferometric Techniques for the Study of Musical Instruments 251
Efthimios Bakarezos, Yannis Orphanos, Evaggelos Kaselouris, Vasilios Dimitriou, Michael Tatarakis and Nektarios A. Papadogiannis

Shock Wave Characteristics in the Initial Transient of an Organ Pipe 269
Jost Leonhardt Fischer

Computed Tomography as a Tool for Archiving Ethnomusicological Objects 305
Sebastian Kirsch

3D Imaging of Musical Instruments: Methods and Applications 321
Niko Plath

How to Interprete Early Recordings? Artefacts and Resonances in Recording and Reproduction of Singing Voices 335
Malte Kob and Tobias A. Weege

Introduction

Big data analysis (BDA) as well as artificial intelligence (AI) are about to enter Phonogram Archives only just like with Telemata or COMSAR. The advances in music information retrieval (MIR) make it possible to analyze musical pieces automatically in terms of rhythm, melody, timbre, or form with a high degree of precision. Intelligent search engines based on AI and deep learning are able to cope with the huge amount of musical pieces recorded and stored in Phonogram Archives over more than about 100 years, tasks which can no longer be performed by researchers 'by hand'. Additionally, such search engines meet the requirements of giving overviews of music from around the world and point to styles, ethnic groups, or traditions which might not be in the focus of researchers and people interested in music at first. Musicians can much easier access music from the past, as well as contemporary developments, by similarity searches rather than by chance.

Universalities in music, as aimed for with such a system, is therefore taken as self-organization. The rules governing music from all around the world appear as emerging properties of the complex interplay of all musical parameters. A rhythm or groove crucially depends on the timbre of the instruments used, a tonal system on the vibrato and pitch glides present, and a timbre on the speed a piece is played. Therefore, the single findings are combined not by simple add-on rules but as interactions of all parameters.

Yet another advantage is the lack of bias though the researcher, like Western approaches to music analysis or the preference of single ethnic groups motivated by nationalistic or political interests. Musical features might be found to be similar in musical styles from ethnic groups which are far apart geographically, linguistically, or religiously. On the other hand, specific features only found in single musical styles would also appear within the same framework making it possible to decide on such ground about how special or common these styles are.

The AI approach has also the advantage to work on multiple levels. By comparing at best all possible musical styles from around the world, a global map can be composed and musical pieces can be compared on such a global level. Still in some cases, it might be interesting to go into more detail on a lower level,

comparing the music of single continents, music traditions, or on performance styles, like singing, instrumental, ritual, electronic, or art music. On micro-levels, the styles of single instrumentalists or composers can be compared.

As metadata are most often included in Phonogram Archives—and traditionally this is the only data available next to the music—sociological or emotional content or symbolic value of the music can be included in such analysis too and compared to the results from the MIR analysis. Here, the expertise of fieldworkers needs to come in to analyze and interpret the results of the algorithms.

Today's search engines of any kind lack of such a self-organizing nature and are therefore subject to ethic and political issues. Engines based on positive evaluations of users on search items (likes, ratings, verbal comments, etc.) easily produce so-called echo-chambers, where only those items are found which are the most popular. Then these chambers exclude items which might be of high quality to many people. Also when suggesting single items, users might be driven by political or economic interests rather than by a scientifically driven evaluation process which are not based on ratings, likes, or suggestions by others. When displaying all found items in global maps, as suggested by algorithms discussed in this volume, users have the overview of what is there at all.

This field of Computational Phonogram Archiving is highly interdisciplinary as it covers ethnomusicology with classical field working and analysis 'by hand,' computer science with focus on algorithms and programming, Internet-based platform programming with all issues of accessibility, networking, responsiveness and data-management, music psychology deciding about the use and content of algorithms and validating their success rate or musical acoustics and physical modeling of musical instruments as contemporary representation of organology. The present volume tries to cover all aspects with grounding ideas and concepts, examples of working systems as well as formulation of problems and future work.

All papers were presented at the first International Symposium on Computational Ethnomusicological Archiving (ISCEA) at the Institute of Systematic Musicology, University of Hamburg, December 7–10, 2017.[1] Everything new has a magic in itself, and having all these people working on different aspects together on this issue for the first time really felt like this! So this issue is also a thank you to all participating. Still its main focus is to present the framework of a future technology implementing the core idea of Systematic Musicology, namely the search for general rules and universals in music, to realize what we all have in common rather than what splits us apart.

In Chapter "Computational Music Archiving as Physical Culture Theory," **Rolf Bader** gives a general overview of Computational Phonogram Archiving, describing the methods of MIR and AI algorithms. After comparing basic system concepts like taxonomies, self-organization or bottom-up/top-down ideas, the chapter describes the approach in a broader sense of a Physical Culture Theory. A Phonogram Archive Standard is proposed which is implemented in the COMSAR system. Adding additional techniques like physical modeling and

[1] The workshop was funded by the Volkswagen Stiftung.

microphone arrays, the standard is widened including state-of-the-art measurement and auralization techniques.

Leslie Tilley starts the section on ethnomusicological fieldwork and archives in Chapter ""The *Lanang* Is the Bus Driver": Intersections of Ethnography and Music Analysis in a Study of Balinese *Arja* Drumming" with a detailed analysis of Balinese *Arja* Drumming. In extensive fieldwork recording and interviewing the most respected *kendang arja* musicians of Bali, the highly complex interlocking patterns appearing when two drummers perform simultaneously are analyzed in great detail. As these patterns are improvised, strategies and rules for a successful performance are discussed. The approach is based on Analytical Ethnomusicology, which is based on fieldworkers' analysis of traditional music.

Another such approach is taken in Chapter "Temperament in Tuning Systems of Southeast Asia and Ancient India", where **Rolf Bader** analyzes several tonal systems collected during fieldworks in Myanmar, Cambodia, Sri Lanka, and Bali. There is a long-going discussion in ethnomusicology about why these tonal systems are so different from Western tunings and differ so much between styles and even within styles. He suggests that the tunings are temperaments in the sense of compromises, just like Western scales, e.g., the equal or the Werkmeister tuning, are compromises between an ideal and constraints making them seem arbitrary at first.

Jukka Louhivuori in Chapter "John Blacking Revisited—Comparative Analysis of Venda Tshikone Dance (1958 and 2009)" discusses the fieldwork and findings of John Blacking, who was much influential in ethnomusicology in the twentieth century mainly by his book 'How musical is man'. There he studied South African Venda music and discussed the stability and variability of songs changing over the years. He also addresses issues of music psychology and cognitive psychology and suggests them to be a crucial part of the analysis of music. The chapter also discusses the relation between music and dance.

One of the largest and most influential Phonogram Archives is the Smithsonian collection and their releases on Folkway records in the twentieth century. **Jeff Place** in Chapter "Smithsonian Folkways and the Associated Ralph Rinzler Folklife Archives and Collections" gives a historical overview of the foundation of the archive and major figures like Moses Ash, who saw Phonogram Archives as education and therefore as stabilizing democracy. Important recordings of Woody Guthry or Lead Belly, recording and storing devices changing over the decades as well as the state of the archive today give an insight in why and how archives come into place, which challenges they face and how they are perceived by the public.

The section is closed by a study on tuning systems of *gamelan* orchestras and their perception by **Gerrit Wendt** and **Rolf Bader**. Recording single instruments. pitch-shifting them, and building a *gamelan* piece by a sequencer software, several versions of this piece were constructed in different tunings, like slendro, just intonation, or Arabian scales. In a listening test asking players of *gamelan* from around the world, it appeared that the constructed pieces were perceived as more or less rough or out of tune which correlates highly with a roughness algorithm, except for the original tuning which is different. This suggests a *gamelan*-specific perception, which causes are not known yet.

After discussing examples of classical ethnomusicological analysis with several tools, the next section gives examples of AI computational tools for approaching a Computational Phonogram Standard. **Michael Blaß** and **Rolf Bader** discuss Hidden Markov models (HMM) and self-organizing Kohonen maps as tools to understand and organize a large set of musical pieces automatically. The tools end in a two-dimensional map displaying similarities of pieces in terms of a musical parameter, here the rhythm of pieces, where fingerprints were calculated by the HMM which are sorted in the Kohonen map. This allows an organization of pieces without the bias of a researcher.

In Chapter "Spatial Manipulation of Musical Sound: Informed Source Separation and Respatialization," **Sylvain Marchand** shows a method of extracting single tracks of a multi-track recording from a stereo audio file. The method is informed such that starting from the multi-track recording a stereo track is created incorporating metadata in the soundfile making a source separation possible which perfectly reconstructs the original single audio tracks. As uninformed sound source separation algorithms still have large errors, at present this is the state-of-the-art algorithm to extract tracks from a given recording. Then new mixes can be made and researchers can concentrate on single tracks. This is an especially useful tool for Phonogram Archives as it allows the storage of a multi-track recording often made in the field, mixed down to a single stereo track.

The Chapter "Finding Music in Music Data: A Summary of the DaCaRyH Project" by **Oded Ben-Tal, Bob L. Sturm, Elio Quinton, Josephine Simonnot** and **Aurelie Helmlinger** tries to bridge the gap between ethnomusicology and computer science. It is based on Telemata, a computer system, implemented at the CNRS Mus´ee de l'Homme in Paris with a collection of ethnomusicological recordings of over 43,000 recordings since 1900. The study analyzes Calypso steel bands in terms of tempo variations and articulations. There is also a creative side of the project where analyzed tunes are recombined into new music presented in concerts.

The interaction of musicians performing live over the Internet is still a challenge faced in Chapter "Experimental Investigations and Future Possibilities in Network-Mediated Folk Music Performance" by **Chrisoula Alexandraki**. It uses the DIAMOUSES software to perform networked music performance (NMP) which allows the instantaneous ensemble performance of several musicians at very different locations in the world. After a detailed overview of the literature, in an experiment it uses Crete folk music with three musicians interacting. Different settings are tested and analyzed. It appears that NMP is possible with the application and its feasibility depending on the settings is discussed.

The section concludes with a discussion on the implementation of Computational Phonogram Archives on platforms of **Jonas Franke** in Chapter "Requirements and Use Cases for Digital Sound Archives in Ethnomusicology". From expert interviews, the needs and demands of such an archive are extracted. Then the challenges and possibilities are discussed both in terms of the interaction of ethnomusicologists and computer scientists, as well as the technical aspects, like the treatment of metadata, MIR algorithms and output. A description of the

COMSAR project in terms of the structure of user access and server implementation is given.

In the last section of this volume possible extensions of Phonogram Archives are discussed, necessary to meet the state-of-the-art of advancements in musical instrument measurement and physical modeling. Chapter "Laser-Based Interferometric Techniques for the Study of Musical Instruments" of **Efthimios Bakarezos, Yannis Orphanos, Evaggelos Kaselouris, Vasilios Dimitriou, Michael Tatarakis** and **Nektarios A. Papadogiannis** shows laser interferometry measurements of musical instruments, here the Crete lyra. The paper gives an overview of the method and shows examples of lyra eigenmodes. The measurements were made with a novel approach of a vibrating mirror reducing measurement noise tremendously.

In Chapter "Shock Wave Characteristics in the Initial Transient of an Organ Pipe", **Jost Leonhardt Fischer** discusses shock waves in flutes as present during the initial transient phase of tone production. He measures these waves in a flute and finds strong correlations with a compressible Navier–Stokes physical model of the instrument calculated on the computer. Such shock waves are important for the establishment of a steady sound and only present during the initial transient phase. The fluid dynamical physical model is therefore able to analyze musical instruments with a great precision and its use in a Phonogram Archive is discussed.

In the field of Digital Humanities in Chapter "Computed Tomography as a Tool for Archiving Ethnomusicological Objects", **Sebastian Kirsch** gives an overview of a project of high-resolution 3D Computer Tomography (CT) of musical instruments. This method allows the scanning of instruments with a high precision which is important for storing and restoration of instruments for museums or as inspiration of instrument builders having the precise geometry of instruments available otherwise not accessible to them. Also the geometries can be used as input to physical modeling to auralize the instruments.

Also in this field are other methods of 3D imaging of musical instruments shown in Chapter "3D Imaging of Musical Instruments: Methods and Applications" by **Niko Plath**. Photogrammetry, 3D surface scanning as well as X-Ray Computer Tomography are discussed as three possible ways to get the geometry of a musical instrument. The challenge to transform these geometries into objects which can be used as input to physical modeling is big, and possible workflows are discussed. Several examples are given and future projects in this field show a great demand both for museums as well as for Phonogram Archives.

Finally in Chapter "How to Interprete Early Recordings? Artefacts and Resonances in Recording and Reproduction of Singing Voices", **Malte Kob** and **Tobias A. Weege** analyzes horns, ducts, and soundboxes as used to record voices on gramophone discs. He shows the transfer functions of several horns and discusses its influence on the recording as well as the singing techniques used with gramophones. It might be the case that singers for such recordings were selected in terms of how good they will sound after their voice was filtered by such a recording horn and how recording engineers were experts in choosing the right horn for different singers.

The volume is the start for the development and establishment of a Phonogram Archive Standard which might be also interesting to music labels or streaming platforms. The aim is to have an intelligent search engine which allows users to maintain an overview of music from all over the world by not only staying in their echo chamber of suggestions of fellows but to enlarge their knowledge and to enhance a deeper understanding on the similarities—and differences—of musical styles from all around the world.

We thank the Deutsche Forschungsgemeinschaft for funding the Computational Music and Sound Archive (COMSAR) project and the Volkswagenstiftung for funding the First International Symposium on Computational Ethnomusicological Archiving (ISCEA) at the Institute of Systematic Musicology, Hamburg, Germany.

Part I
Overview

Computational Music Archiving as Physical Culture Theory

Rolf Bader

Abstract The framework of the Computational Music and Sound Archive (COMSAR) is discussed. The aim is to analyze and sort musical pieces of music from all over the world with computational tools. Its analysis is based on Music Information Retrieval (MIR) tools, the sorting algorithms used are Hidden-Markov models and self-organazing Kohonen maps (SOM). Different kinds of systematizations like taxonomies, self-organazing systems as well as bottom-up methods with physiological motivation are discussed, next to the basic signal-processing algorithms. Further implementations include musical instrument geometries with their radiation characteristics as measured by microphone arrays, as well as the vibrational reconstruction of these instruments using physical modeling. Practically the aim is a search engine for music which is based on musical parameters like pitch, rhythm, tonality, form or timbre using methods close to neuronal and physiological mechanisms. Still the concept also suggests a culture theory based on physical mechanisms and parameters, and therefore omits speculation and theoretical overload.

1 Background

1.1 Kinds of Music Systems

The aim of Systematic Musicology, right from the start around 1900, is the seek for universals in music, for rules, relations, systems or interactions holding for all musical styles of all ethnic groups and cultures around the world [1]. Music recordings and Phonogram Archives played a crucial role for establishing the field, as only after the invention of the Edison phonograph for recording music on wax cylinders [2] it was possible to compare music recorded by ethnomusicologists. The first of such archives was the Berlin Phonogram Archive established by Erich von Hornbostel and Carl Stumpf with a historical recording of a Thai *phi pha* orchestra at a visit

R. Bader (✉)
Institute of Systematic Musicology, University of Hamburg, Neue Rabenstr. 13, 20354 Hamburg, Germany
e-mail: R_Bader@t-online.de

R. Bader (ed.), *Computational Phonogram Archiving*, Current Research in Systematic Musicology 5, https://doi.org/10.1007/978-3-030-02695-0_1

in the Berlin Tiergarten 1900. Many recordings followed, like those of Jaap Kunst recording music of Indonesia [3].

1.1.1 Taxonomy

One way of giving the endless variety of musics a system was the Hornbostel/Sachs classification of musical instruments [4] in align with the Sachs dictionary of musical instruments [5] as a detailed basis for such a classification. This classification system still holds today and is only enlarged by electronic musical instruments, whos development started in the second half of the 19th and was very prominent in the second half of the 20th century [6]. The success of this classification is caused by its classification idea of sorting musical instruments according to acoustical properties, namely their driving mechanisms. So instruments can be plucked, bowed, blown or struck, which produces similar timbres. This classification system is therefore showing a unity in the variety by comparing instruments. Therefore the field of Systematic Musicology in these days was called Comparative Musicology [7]. Taxonomies have been applied in many contexts in the field. A prominent example is that of musical styles forming feature lists [8] as applied to singing styles around the world [9].

1.1.2 Self-Organization

Classification is a hierarchical structure with global nodes followed by subnodes to differentiate plucked or blown instruments into many subcategories. This is only one kind of a system, and several others have been proposed. One such way is describing music as a self-organized system. The sounds produced by musical instruments are showing a very simple behaviour, the harmonicity of their overtone structures, only due to a very complex system of linear and nonlinear substructures, like turbulence in wind instruments or the bow-string interaction of the violin family [10]. The perception of music in the brain in neural networks is also a self-organizing process [11–14], where many neurons interact in nonlinear ways to result in simple outputs, the perception of timbre, rhythm, melodies or musical form. Therefore self-organization is a second system proposed to understand musics from all around the world in a more differentiated way [10].

1.1.3 Psychoacoustic Bottom-Up Process

Yet a third kind of system to understand music is often used in Musical Signal Processing and Music Information Retrieval (MIR) [15–17]. Here algorithms investigate the digital waveform of recorded music to retrieve information from this waveform, such as pitch, timbre, rhythm or other psychoacoustic parameters like roughness,

brightness, density or the like. Classification of musical instruments is performed, as well as many other tasks like score following or Networked Music Performance [18]. Here the computer understands the music and can tell a player where he is in the score. Musicians around the world play together over the internet, and the task of the computer is to synchronize their playing, working around the restrictions of the speed of light, delaying the transmissions. This can be achieved by estimating the played music of a musician before it is actually played. Piano-roll extraction is yet another such task, where the computer understands typically piano music and prints a score from the sound wave file [19].

1.2 Basic Types of Algorithms

Many of those tasks are realized by algorithms estimating a bottom-up approach to music retrieval, where the sound file is analyzed in terms of its spectrum or cepstrum at first which is then further processed using more complex algorithms to end up in the retrieval result [17]. To which extend those algorithms represent perception of music by humans, the processes physiologically present from the cochlear and the neural nuclei following up to the primary auditory cortex and beyond is not too much discussed. Indeed neural processing is self-organization and therefore many of the algorithms are often only roughly related to perception [20–22]. Still this is not the main aim of such algorithms which can be seen as engineering solutions to a given task and often perform very well.

Other more complex algorithms have been proposed coming closer to human perception, like self-organizing maps of the Kohonen map type [11–14] (see also Chapter "Content-Based Music Retrieval and Visualization System for Ethnomusicological Music Archives" of this volume). Here features of musical sounds are sorted in a map according to their similarities. Still as the map is organizing itself, no initial estimations are needed to decide about the way similarity may be measured, the features of the different sounds decide about this on their own. Also algorithms estimating the fractal dimension of sounds show considerable relatedness both to the processing of sound in humans as well as the perceptional sensation of musical density as a simple result to a complex computation [10].

Another kind of algorithms used for such tasks are Hidden Markov Models (HMM) [23]. Here the temporal development of events are predicted as the result of some hidden process. This process consists of the transition between a small amount of states, like musical pitches or musical instruments. The development of their appearance is modeled here as the probability of the transition of one state switching into another state, so e.g. one pitch followed by another one. As such transition probabilities are of statistical nature likelihoods describe the process and therefore the output is not fixed beforehand leaving space for arbitrariness. Still to which extend these models fit human perception is under debate (for details see Chapter "Content-Based Music Retrieval and Visualization System for Ethnomusicological Music Archives" in this volume).

Yet a totally different kind of algorithm describes music production of musical instruments. Physical Modeling is a set of methods to produce the sound of instruments by knowledge about the instrument geometries and the physical laws governing their vibration [24]. Several stages of complexity and simplicity exist here, from lumped models, digital waveguides or delay lines [25] to whole body geometries solving the differential equations governing the vibrations of plates, membranes or turbulent air flow [10, 26] (see also Chapter "ShockWave Characteristics in the Initial Transient of an Organ Pipe" in this volume). These algorithms use the detailed geometry and solve the problems in a time-dependent manner, resulting in very realistic sounds and estimations of vibrating frequencies, transients or radiation. Using extreme parallel computation on an Field-Programmable Gate Array (FPGA) these geometries can be simulated in real-time [27, 28]. Understanding the behaviour of instruments is often achieved in a bottom-up way by trying different models, adding or leaving out geometrical details, and from the comparison between the computed and measured sound decide about how the instrument works in detail. Other more simple models are able to show relations between musical instruments and instrument families more easily by starting with global estimations and adding necessary features in a top-down way [10].

Measurements of musical instruments are also a crucial part of the understanding the instruments and their relation one to another. Whole body measurement techniques, like microphone arrays [29], laser interferometry [30] (see also Chapter "Laser-Based Interferometric Techniques for the Study of Musical Instruments" in this volume) or modal analysis [31] all give a detailed picture about the spatial and temporal development of the vibrations and transients of musical instruments. High-speed cameras and sub-pixel tracking analysis show the movement of strings or reeds [32]. This understanding leads to estimations of the global behaviour of instrument radiation, the role of different instrument parts in sound production, the use of materials like woods, hybrid- or metamaterials, the interaction of musical instruments with room acoustics or between instruments and players. Modern high-resolution methods of computer tomography (CT) give very detailed geometries of the instruments (see Chapters "Computed Tomography as a Tool for Archiving Ethnomusicological Objects" and "3D Imaging of Musical Instruments: Methods and Applications" in this volume) which can be used as input to physical modeling, showing details within the structures not accessible from their surfaces, or giving estimations of material parameters like density or Young's modulus. Therefore these methods are able to compare instruments and instrument families and give insight into building strategies and methods.

The different approaches to understanding music should at best all be used in a Computational Music and Sound Archive (COMSAR) as proposed here. Traditional phonogram archives only consist of recordings and their metadata, like the country they have been recorded, the musicians playing, the instruments in use etc. [33–37]. Still to address the aim of Systematic Musicology of finding universals in music [38], understanding its system, its production and perception need to use the analysis and analysis-by-synthesis tools discussed above. Including all these tasks is a tremendous effort, still all these fields nowadays show a high degree of specialization and are

able to give detailed and robust results. Therefore to combine them together to an automatic analysis and search engine is straightforward.

1.3 COMSAR as Big Data Solution

Such automatic systems are needed in many fields. The endless amount of accessible music recordings via the internet, on CDs and in archives makes it practically impossible for a single researcher and even for research teams to perform these analysis by hand. Such a Big Data problem needs automatic tools for researchers to cope with. Additionally, the amount of methods and their complexity are so large that it is not feasible to have one researcher perform all tasks.

Also in terms of the 'buzz' the internet produces in terms of the endless variety of musics, research and consumer demands, search engines are needed to point researchers, musicians, listeners and music lovers to music they would hardly find otherwise. Such search engines need to be based on real musical parameters. Existing search engines for music used for mood regulation or work-outs sort music mainly in terms of its tempo and on its vitalizing properties. Still the musical styles of the world are so many and so differentiated that such simple parameters are not able to represent traditional music, all-in-a-box productions of musicians of ethnic groups in remote areas, boy groups in jungle regions combining their tradition with Western harmony or electronic music, free-improvisation, global Hip-Hop or Electronic Dance Music. Here much more differentiated algorithms are needed, and all of those mentioned above should be combined.

1.4 COMSAR as Physical Culture Theory

The COMSAR standard is not only to update traditional phonogram archives with modern methods and algorithms and not only about coping with Big Data. It is also the attempt to realize a culture theory based on physical reasoning.

As has been shown, both, musical instruments as well as neural brain networks are self-organizing systems [10]. They both are highly nonlinear and intensely coupled only to output a very simple behaviour. In terms of musical instruments this output is the harmonic overtone spectrum which would not be perfect without the self-organizing process at all. In terms of music perception and production, pitch, rhythm, melodies, musical form and other features are the results of self-organization and synchronization in the human brain.

Historically musical instruments have been developed and built by humans over at least the last forty thousand years according to the physiological and physical mechanisms the human brain and body is built of. The human voice, as well as animal vocalization are also self-organizing processes. They produce sounds which are not often found in non-living systems, namely the harmonicity of the partials.

As voice is meant for communication, harmonic sounds are evolutionarily related to semantics. Therefore the semantics found in music need to be there because it is built-in the human auditory system reacting to harmonicity of the sounds.

Self-organization is the base of life compared to dead matter, it is turning non-living things to life. Its main issues are maintainance of life in a destructive world, differentiation in parts to fulfill difficult tasks and the ability to assimilate in a changing environment. So building musical instruments as self-organizing systems means to make them similar to living systems. They exhibit behaviour we know only from animals or humans, harmonic tone production. Musical instrument builders have obviously decided to make this physical feature the core of musical instruments, and therefore the core of music as a cultural phenomenon. So the core of musical instruments is their self-organizing nature. We have built a music culture as artificial life by building musical instruments and perform on them.

A Physical Culture Theory is taking culture as a physical and physiological self-organization process building artificial life and therefore extending our life by inventing physical tools and processes which again work as self-organizing systems. Therefore the culture we build appears to us as a living system. The music speaks to us, the development of musical styles follow living behaviour, styles are born, live, die and are remembered, become legend. Musicians fuse with their instruments and experience them as having their own live, relate to them very similar to the way they relate to humans.

COMSAR, as implementation of many of the mechanisms and systems is therefore approaching music as a living culture, a self-organizing process. Of course it is only an approach yet, still extending the system in the future towards more and more precise algorithms and tools is only a matter of time.

Due to the difference of the algorithms discussed above, in the follow we give a deeper insight into main properties and research done in these fields, without being able to mention all of the works done here. MIR is already implemented in Phonogram Archives, physical modeling, microphone array techniques are not, and therefore work in ethnomusicology using these tools are given. Self-organizing maps as front end for search engines are discussed.

2 Tools and Applications for COMSAR

In 1978 Halmos, Kszegi, and Mandler coined the term Computational Ethnomusicology for using MIR also in regard to non-Western music [39]. Since then many tools and applications have been suggested to retrieve, sort and understand musical content from sound.

2.1 Bottom-Up MIR Tools

In early attempts to extract musical parameters from sound, simple multi-line textures consisting of two voices were considered, which must not have overlapping overtones [40]. More modern approaches include percussion sounds, meter and rhythm estimation [41] and are designed for the analysis of harmonic as well as percussive instruments [42], including psychoacoustic knowledge [43], or using different approaches in the matrix domain [44]. Similarity matrices of spectral features, like Mel Frequency Cepstral Coefficients (MFCC), amongst others, have also been proposed to relate parts of a piece, like verse or chorus [45]. Singular Value Decomposition (SVD) has also been employed in this context [46]. Recently TARSOS, a platform to extract pitch information from sound using 1200 cent per octave has been developed [47] to suit demands in Computational Ethnomusicology [39].

The first task to accomplish in analysing audio signal will invariably be the detection of the onset of any given signal event (see [48] for a review). The approaches employed here range from measurement of strong amplitude raise and phase differences, to fluctuation estimation. It appears that the choice of the onset detection algorithm depends on the type of sound to be analyzed. For percussive sounds, measurement of amplitude raise is sufficient, yet for fusing tones, like piano or violin sounds, measurement of fluctuations seems more promising. As fluctuations on a phase level seem to include both to a great extend, the approach favoured by the applicant and his staff is therefore based on a Modified Modulation Algorithm [18].

A second task is the estimation of pitch, often referred to as f0-estimation, i.e. detecting the fundamental partials of a given harmonic spectrum [17, 49]. An approach used in the context stated in this proposal is based on algorithms like Autocorrelation Functions, but furthermore employs Correlogram Representation for f0 estimation in multi-line textures [50, 51]. This robust method allows estimation of harmonic overtone structures within very short time frames. Additionally, to estimate if a piece is single- or multi-line, the Fractal Correlation Dimension is appropriate, as the integer dimension number thus obtained constitutes the amount of harmonic overtone series present in a given musical sound [10, 52].

2.2 Self-Organizing Tools

Representation of the results of an analysis for use in IR algorithms has been proposed in several ways. COMSAR uses self-organizing maps (SOM), Hidden-Markov Models (HMM), and correlation matrices, all based on the extracted data.

Self-organizing Kohonen maps have been proposed for pitch and chord mapping [12, 53], and for sound level assessment [11], for a review see [10]. This method has also been successfully applied to related fields, such as speech estimation [54], and soundscape mapping. Here the feature vector extracted by the MIR algorithms, consisting of pitch contour, spectral centroid, fluctuations, inharmonicity, etc., is fed

into an Artificial Neural Network within a defined training space. After the training process, any such system should be able to identify new feature vectors by itself and will therefore be able to define a parameter space for these features for all of the analysed archival assets, and will be able to detect structural similarities on a best-estimation basis.

2.3 Hidden-Markov Model (HMM)

A complementary approach to be employed is the implementation of the Hidden-Markov-Model (HMM), used for stochastically estimating transition probabilities between hidden states, which, performed consecutively, results in an event series, as present with both, musical rhythm and melody. These models have been used extensively for musical applications [18, 55]. The Markov model consists of musically meaningful states. So when representing, for example, a multi-line rhythm, these states could be bass drum, snare drum, hi-hat, tom-tom, etc. These are mathematically represented as a Mixed Poisson distribution.

Additionally, a transition matrix between these hidden states will be calculated using an Estimation-Maximization (EM) algorithm [23]. Both, the Poisson distribution and the transition matrix determine the musical parameters, rhythm, melody and texture. This representation may then be compared to all previously analysed assets in the ESRA database, again forming a state-space, detecting similarities, relate objects, etc.

3 Architecture of COMSAR

A MIR-based data infrastructure and classification scheme is to be implemented within the framework of the ESRA database currently under development to be able to categorize the database content in regard to basic musical parameters derived from the digital audio data stream.

The three main musical parameters which are treated using the MIR analysis described in this proposal are pitch (melody, texture), rhythm (single- and multi-line), and timbre (single- and multi-line). The MIR structure has two main threads, the timbre thread (TT) and the pitch thread (PT). As TT deals always with the whole sound information, PT performs a pitch extraction from the sounds and proceeds with pitch information only.

3.1 Timbre Thread (TT)

The first step in TT is a segmentation of the audio file in terms of onset detection (OD). Here, two main methods are used, the fluctuation method for fusing tones [18] and a simple amplitude model for percussive onsets [17]. From the segments three MIR estimations are performed: a Timbre Thread Rhythm (TTR), a Timbre Thread Timbre Multi-line (TTM) and a Timbre Thread Timbre Single-line (TTS).

3.1.1 Timbre Thread Rhythm (TTS)

TTR does take the sound played by several instruments as one; it does not attempt any splitting of compound sounds into individual instrument sounds. As discussed above, retrieving individual sounds of musical instruments from a multi-instrumental recording is theoretically impossible, because of the fact that no clear association with all partials of harmonic pitch structures can be assumed from the sound alone without any further knowledge. Still to be able to deal with more complex rhythms, in the PT section (see below) a multi-line estimation is performed to detect the most probable events without the need to extract the sounds perfectly.

Within the TTS, for each segment a spectral centroid is calculated as the most prominent parameter of timbre perception. The list of centroid values of the onsets found is then fed into a Hidden-Markov Model (HMM), using a Poisson Mixture Model (PMM). The results of the HMM are the parameters of the PMM, which represent the rhythmical structure of the centroid values of the onsets. This PMM, as well as the Transition Probabilities (TP) are calculated for all objects in the database and a correlation matrix between all PMMs and TPs is calculated to relate the different rhythm PMM structures in terms of similarity.

3.1.2 Timbre Thread Timbre Multi-line (TTM)

As discussed in the Pitch Thread (PT) section below, it is estimated if a given recording contains multi-line or single-line melody (this may also be judged aurally and used as additional, external input). Additionally, the fractal correlation dimension D of a given piece is calculated for adjacent sound sections of 50 ms. If 100 ms after the initial transient $D \geq 2.0$, the sound has more than one harmonic overtone structure and therefore is considered multi-line. Within this definition, all percussion objects are multi-line, too. This is reasonable also if only one drum is played. If the piece is found to be multi-line in nature, the TPM algorithm estimates a feature vector of each segment provided by the onset detector, using spectral centroid, fluctuations within the steady-state of the sound, amount of chaoticity of the initial transient, and other related features found with timbre perception of multi-dimensional scaling events. These features are calculated for adjacent times within each segment to end in a multidimensional trajectory of the sound development, as found crucial to explain

nearly endless possible sounds within a low-dimensional timbre space, by adding the temporal development of the sound within this space.

This feature vector is then used as input to a self-organizing Kohonen map. After training, this map constitutes a two-dimensional representation of the objects in the database. All segments of all objects are then fed into the map, where the neuron with maximum similarity between the given segment and this neuron positions the segment within the map. Therefore, segments or objects can be estimated for similarity from the trained map.

3.1.3 Timbre Thread Timbre Single-line (TTS)

If a piece is found to be single-line all through, as discussed above, the same procedure is performed, training a Kohonen map with the feature vector of the given sound. Again the trained map is then able to relate all segments and all pieces, and give similarity judgements. The reason why the single- or multi-line cases are separated is to have one map which is able to classify single instruments alone, while the other is able to deal with orchestrated multi-instrument sounds. So if a musical instrument is to be judged in terms of similarity, the TTS can be used. Another reason is the problem of dealing with the different pitches of the sounds. The TTS map will classify both, pitch and timbre. As pitch is the most prominent factor in musical instrument similarity judgements, the map will have different regions for different pitches. Then within each region the differentiation in terms of timbre is present. This is automatically performed by the map. Still it is necessary, as one instrument may sound considerably different within different registers. Differences in articulation within one pitch region will again be met by the differentiation of the map within the pitch region of the sound investigated. This cannot be done with multi-line sounds, as here virtually endless possibilities of pitch combinations can be present.

3.2 Pitch Thread (PT)

PT is representing a piece on the score level, although of course it also needs to start from the recorded sound. So, first PT performs a pitch extraction, both single- and multi-line. Two main algorithms are used here, the correlogram for multi-line and the autocorrelation function for single-line sounds. The correlogram is detecting whole overtone structures, which are related to pitch, and finds the basic frequency for it within small time frames of about 20 ms. As it also displays multiple harmonic series, the pitches of different instruments can be detected with high frequency resolutions. If a piece is single-line, this algorithm can be used, too. Additionally, an autocorrelation estimation of small time frames of again about 20 ms is performed, adding information to the correlogram. The result is a temporally and spectrally high resolution function of the harmonic overtone series, the pitches, over time, for the whole object. From this pitch texture, again three musical parameters are calculated, the

Pitch Thread Rhythm (PTR), the Pitch Thread Texture (PTT), and the Pitch Thread Melisma (PTM).

3.2.1 Pitch Thread Rhythm (PTR)

For PTR, from the pitch texture, note onsets are calculated to end up in a musical score. This score consists of pitch events, both in terms of Western pitch classes as well as in term of their microtonal precisions up to 1200 cent per octave over time. As with the TTR, a Hidden-Markov Model (HMM) is used with a Poisson Mixture Model (PMM) which has as many hidden states as are different pitches appearing in the given object. The PMM and the Transition Probability (TP) calculated by the HMM then represents the objects, which can therefore be related in terms of similarity.

3.2.2 Pitch Thread Texture (PTT)

Here, from the pitch texture again the onsets are calculated, again to end in a score. For PTT, this score itself is used to correlate the objects in terms of their similarities. PTT is meant for objects with no glissando, where the pitch texture holds the main information of the object.

3.2.3 Pitch Thread Melisma (PTM)

Contrary, PTM is meant for objects in which pitch changes are very important, in terms of vibrato, glissandi, etc. Again after performing onset detection, the objects are divided into small segments for each played note. Still, all detailed pitch information over time within the segments is preserved here. Again, a Kohonen map is used to represent possible ornamentations, melismata, glissandi, or vibrato. So here special ornamentations or melismata can be compared, rather than whole objects themselves.

3.3 Summery of Threads

After performing all these analysis of all objects, six parameters result, which estimate the basic musical parameters timbre, pitch, and rhythm for each piece:

Timbre
Kohonen map of multi-line timbre (TTP) Kohonen map of single-line timbre, musical instrument sounds (TTH).

Pitch

Score for multi- and single-line objects (PTT) Kohonen map of melisma for all segments of all objects (PTM).

Rhythm
Hidden-Markov Model of multi-line fused sounds based on sound level (TTR) Hidden-Markov Model of multi- and single-line objects based on pitch level (PTR).

With these six models all objects within the archive can be compared in terms of all the sub-features present in the models. Also new objects or sounds can be compared to all the existing features in the model. The similarities proposed by the algorithms then need to be judged by listeners and experts.

4 Including Musical Instrument Measurements and Modeling in COMSAR

Organology was part of understanding and systematizing music as part of the Hornbostel/Sachs classification. Musical instrument dictionaries like that of Curt Sachs (see above) mention and describe thousands of musical instruments from all over the world. Features like their origin and use in the musical culture, the material they are built of or the building process are documented. These dictionaries are a useful source when it comes to identifying instruments collected in the field and for giving information about their content.

Still research is way ahead in terms of the acoustics, properties and building processes of musical instruments. The basic principles of how musical instruments vibrate and radiate sound have extensively been studied (for reviews see [56–59]). Many musical instruments have been investigated in great detail, mainly those of the West, but also many others all over the world. The materials used are known in terms of their material parameters like Young's modulus, density or internal damping. The building process of many instruments have been described not only as plain craftsmanship but also in terms of the acoustical and musical function these processes have been motivated by.

Several theoretical frameworks on the acoustical properties of musical instruments have been developed over the last decades making it possible to classify them not only in terms of their driving mechanisms, like Hornbostel/Sachs have done, but by their physical mechanisms and features. The Impulse Pattern Formulation (IPF) considers musical instruments as working with short impulses caused by one part of the instrument, e.g. the force of a string acting on a body or the pressure impulse of wind instruments produced at the players mouth. These impulses are transferred to other parts of the instruments, like top- and back plates, rims and ribs, are filtered and return to its origin. In a self-organizing process this system starts with a transient phase which is complex and chaotic only to organize itself after a short time of maximum 50 ms to end in a harmonic overtone sound radiation [10]. Other proposals are that of a nonlinear driving generator and a linear resonator which interact, producing sound, or that of phase-locking of different partials (for a review see [10]).

Also the geometry of the instruments are known in great detail. High-resolution Computer Tomography (CT) scans of whole instrument geometries display the instruments with resolutions of a fraction of a millimeter (see Chapter "Computed Tomography as a Tool for Archiving Ethnomusicological Objects" in this volume). From these results material properties can be derived, like density, speed of sound, diffraction or internal damping. Therefore the rough estimations of geometrical data have been replaced by detailed and precise measurements.

The radiation of musical instruments have also been measured extensively, as only during the last years technological advances have made it possible to recored single instruments with microphone arrays consisting of up to 128 microphones when recording single sounds and up to several thousand microphone positions when recording multiple sound instances (for a review see [29]). Some techniques allow the back-propagation of the radiated sound to the radiating surface, the musical instrument geometry. This means a measurement of the instrument vibration all over its geometry within and therefore a measurement of the internal vibrational energy distribution, the role of geometrical parts to the acoustical output, or an estimation the radiation of the instrument at any place in a performance space.

Physical modeling of musical instruments have also been performed extensively over the last decade (for a review see [24, 56, 60]). Here the differential equations governing the vibration of the instruments are used with a geometrical model of the instrument to make this virtual instrument vibrate in silico. High time and spatial resolution allows the precise modeling of the instrument and the production of a sound very close to the original sound of the instrument. By changing the mathematical model the role of different kinds of vibrations, couplings or instrument parts can be shown. By changing the geometry or the material parameters the instruments can be understood in terms of why which geometry is used and how changed here would change the sound of the instrument. The use of geometrical changes or alternative materials can be tested before building the instrument, a property needed nowadays as climate change forces new wood species to be planted e.g. in Europe.

All these contemporary features of musical instrument research, models and experimental setups need to be part of a phonogram archive standard in the near future. Two examples are given below, one about the use of microphone arrays to measure the acoustical behaviour of a *lounuet*, a New Ireland friction instrument, and the use of paste to tune the membrane of a Burmese *pat wain*. Both examples show the clarification of ethnomusicological questions about the instruments.

A modern way of using microphone-array measurements, high-resolution CT scans or physical modeling need to be developed in the future and faces several issues, like server space, computational capability of servers or the retrieval of the information necessary to built such models. Still solutions are around for all aspects so that the implementation of these features to COMSAR is mainly a matter of time and energy and not a fundamental problem.

The advantages of such methods are many. Instrument builders could go online and look for the instrument they build, or for similar ones, understand more about its vibration, radiation or the influence of several parts or materials on the sound, include changes of the material or parameters online and listen to the resulting sound of such

a new instrument instantaneously. They could then decide to built such an instrument or thing of different changes. They could also be inspired by similar instruments, their building process, materials or sounds and decide to try new instruments similar to their traditional one.

Researchers could estimate how important different aspects of the instrument for sound production are. Some building processes of instruments are sound decisive, some may be needed in terms of rituals used for the building process, and some might be pure myths, traded by tradition rather than by a sound idle and unnecessary or even unwanted today. In the history of instrument building of Western instruments, like violins, pianos, trumpets or guitars, many of these myths have been identified over the years and from an ethnomusicological standpoint it is important to know which stories are true and which are not.

Also from an educational point of view such a system would be highly attractive to young people interested in the music of the world and used to use the internet, search engines and simple music production systems to be creative. Such tools could be rated as 'cool' and contemporary and therefore be used with the by-effect of making them understand the principles of musical instruments and their use.

In terms of replacing wood and other natural material becoming scarce nowadays with artificial materials, metamaterials or the like, such tools would be highly welcome, too, as everybody could try online how such changes would effect the sound and if they might be used.

Many other applications are to be expected due to an inclusion of these techniques in phonogram archive standards in the near future.

5 Microphone Array Measurement of New Irland *Lounuet* Friction Instrument

This example of the use of a microphone array is about to show that the *lounuet*, a New-Irland friction instrument does only work when the chambers are built in a clever way, that for each plate played they work in combination to resonance the tone. 'Tourist' instruments not taking this into consideration do not sound at all, their radiation is too weak. Only when the chambers are carved correctly the instruments sounds loud.

Friction instruments are driven by a player rubbing over the instruments surface using his hands or fingers, any kind of tissue, or hair. In the Hornbostel/Sachs classification of musical instruments [4, 5] friction instruments belong to the class of 'Streich-Idiophone' where 'Streich' means 'bowing' as well as 'rubbing'. This class is subdivided into rubbing bars and rubbing carillons (*Glockenspiele*). Furthermore, the rubbing bars are divided into transversally-vibrating like the nail violin (see [61] for more details about this instrument) or the nail piano, and longitudinally-vibrating, like the *Clavicylinder* or finger rubbing, and finally into transversal/longitudinal vibrating, where actually the only example is the *Euphone* invented by Chladni. The

carillons are further divided into standing idiophones, glasses and rotating bowels, like e.g. the *armonica*, the glass harmonica invented by Benjamin Franklin in 1755.

In all cases, an idiophone is driven through friction, may it be a bow, a cylinder, or a finger. This is a highly nonlinear process like that of a violin bowing [62, 63]. The 18th century was keen on these instruments, where the *Euphone* of Chladni is a good example. He coupled a glass bar, driven by friction, with a metal bar to indirectly drive it. Although for the glass harmonica compositions exist, e.g. by Mozart, most of these friction instruments were not very popular, mainly because the tone comes in pretty slowly and so fast playing is hard to realize. Still many of them have been invented as evident in the list Sachs gives [5] p. 425:

> Cölison, Euphon, Euphonia, Glasharmonika, Huli-huli, Klavizylinder, Kulepa ganez, melodikon, Melodion, Nagelgeige, Nagelklavier, Panmelodikon, Stockspiel, Terpodion, Trochelon, Uranion, Xylomelodichord, Xylosistron

5.1 The Lounuet *Instrument*

The *lounuet* investigated here is located today at the Überseemuesum in Bremen, Germany (inventory number: D 14488). It is not perfectly clear when it was brought

Fig. 1 The *lounuet* of the Überseemuseum as used in this investigation

to Bremen, the second world war destroyed much of the archive of the museum, but most likely before 1914 [64]. It is one of four instruments there, although only D 14488 can be played, the others are partly damaged. Figure 1 shows the instrument investigated in this paper. It is 51.5 cm long, 19 cm at its maximum width and 23.4 cm in maximum height and built out of one block of *savaf*, *sebáf* (Alstonia villosa), or *bauvît* wood, a very hard wood growing in Papua New Guinea. It has three plates which have been burned out of the block using glowing hard wood. Then the plates are cut out sawing the remaining 'bridges' using a liane. The body is then polished using shark skin. The whole process involves many rituals as the instrument has been highly respected in the society of the tribes in New Ireland and the owner is expected to live a life strictly according to the societies rules. The instrument was often buried with the player [65].

Messner who conducted fieldwork in New Ireland in the late 70th of the 20th century reported that still then instruments have been built, still these were and could no longer be played as they did not work properly anymore.[1] Also the old knowledge of building these instruments was mostly lost. Still he was able to record the older instruments with professional players.

The instrument exists in three sizes, the smallest is 12–20 cm, a medium is 25–35 cm, and the largest, the only one called *lounuet*, is 40–65 cm long. The names of these instruments are derived from birds and frogs and the sounds are supposed to imitate these animals. The smallest has a squeaking or rough sound more like a frog, the medium is tuned higher but much smoother while the *lounuet* is tuned medium and the only one producing a pitched tone (Table 1).

5.2 Chamber Resonances

The three lowest resonance frequencies of the D 14488 *lounuet* which are also the rubbing fundamentals were compared by knocking on all three plates. In all three knocking spectra all three peaks appear, only when knocking on the highest plate this peak is very weak. In Figs. 2, 3, and 4 the radiation patterns of these peaks are shown when knocking on the respective plates compared to the rubbing radiation. The arrows indicate the plate which was knocked on, top to bottom is highest, medium, and lowest. On the right the arrow indicates a rubbing over the respective plate. Clearly, the radiation depends on the knocking point, still the basic patterns which appear when rubbing are preserved.

With the lowest frequency of around 500 Hz in Fig. 2 the pattern is the same when knocking on the lowest and medium plate (although vice versa, still this means the same pattern as the reference point of the phase in a harmonic motion is arbitrary, so phase/anti-phase becomes anti-phase/phase after half a period). Only when knocking on the highest plate the knocking point itself appears. As mentioned above this is the only peak in the spectra which is very weak. This finding supports the above

[1] Messner, personal communication.

Table 1 Fundamental frequencies and cent values of four friction woods as investigated by [66]: (1) Collection Otto Finsch (between 1879 and 1882), Kapsu in Nord-New Ireland; (2) collector unknown; (3, 4) Messner, collected around 1978

Lamella	(1) f(Hz)	(1) cent	(2) f(Hz)	(2) cent	(3) f(Hz)	(3) cent	(4) f(Hz)	(4) cent
Low	510	0	864	0	340	0	480	0
Middle	780	736	1072	373	430	407	520	139
High	1200	1481	1232	614	780	1031	777	834

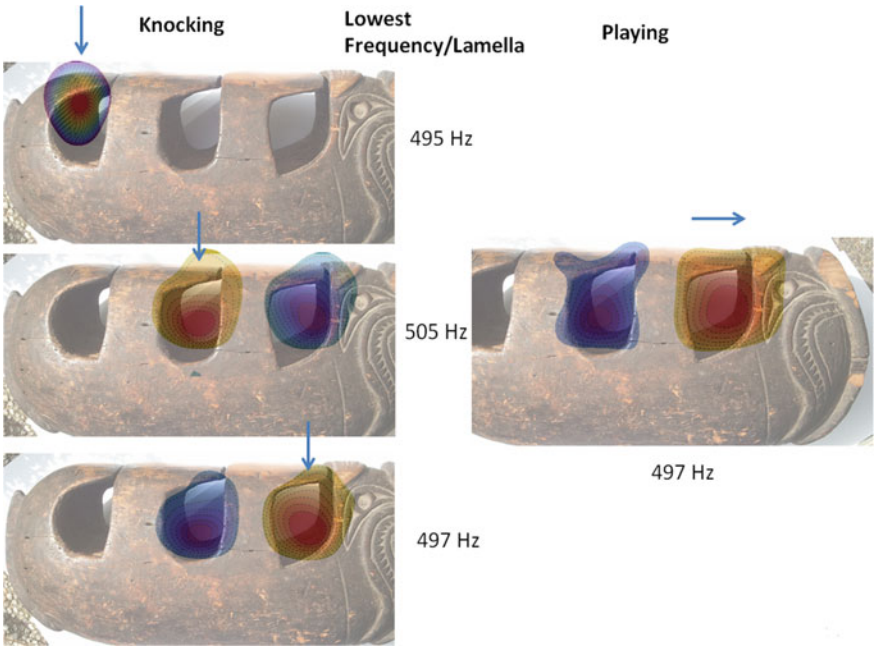


Fig. 2 Comparison between knocking on different plates and playing the lowest plate for the lowest partial in the respective spectra. The dependency between the driving point and the radiation pattern clearly appears. The very strong radiation from the highest plate with this partial comes from the very low energy from this peak at this driving point, in other words, it is nearly not driven. The resonance of this frequency happens within the coupled two lowest chambers

finding that there is a strong coupling between the lowest and medium chamber for the lowest plate frequency but not a coupling to the highest chamber. Also the phase relations between the lowest and medium chamber are the same as for rubbing.

With the medium frequency of around 640 Hz shown in Fig. 3 the same pattern of an anti-node/node/node pattern appears again pointing to a three-chamber coupling between them. The strong anti-node between the medium and the lowest chamber is also present. Still knocking on the highest plate results only in a radiation of the highest and the medium chamber still in the same phase relation. So again the radiation also depends on the driving point which can slightly also be examined when

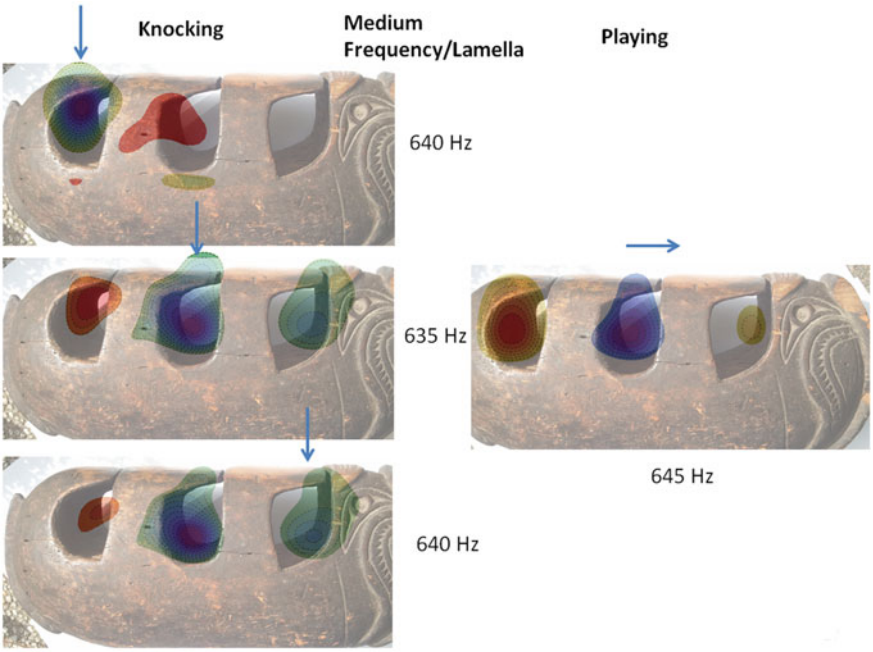


Fig. 3 Comparison between knocking on different plates and playing the lowest plate for the lowest partial in the respective spectra. The dependency between the driving point and the radiation pattern clearly appears. The resonance appears between all three chambers. Although the medium and lowest chamber have nearly the same phase, between them a strong anti-phase appears with nearly no radiation

knocking on the lowest or the medium plate. When knocking on the lowest plate the radiation from the highest chamber is weaker than when knocking on the medium plate.

Figure 4 of the highest frequency around 740 Hz displays right the same behaviour as the previous two frequencies. The anti-node/node/anti-node pattern reappears and the radiation is again depending on the driving point. So again the basic pattern was found as with rubbing.

So the instrument does only sound loud when the resonances appear which are always combinations of the resonance chambers. So to understand the instrument it is necessary to include such measurements in a phonogram archive with the attempt not only to display but to understand.

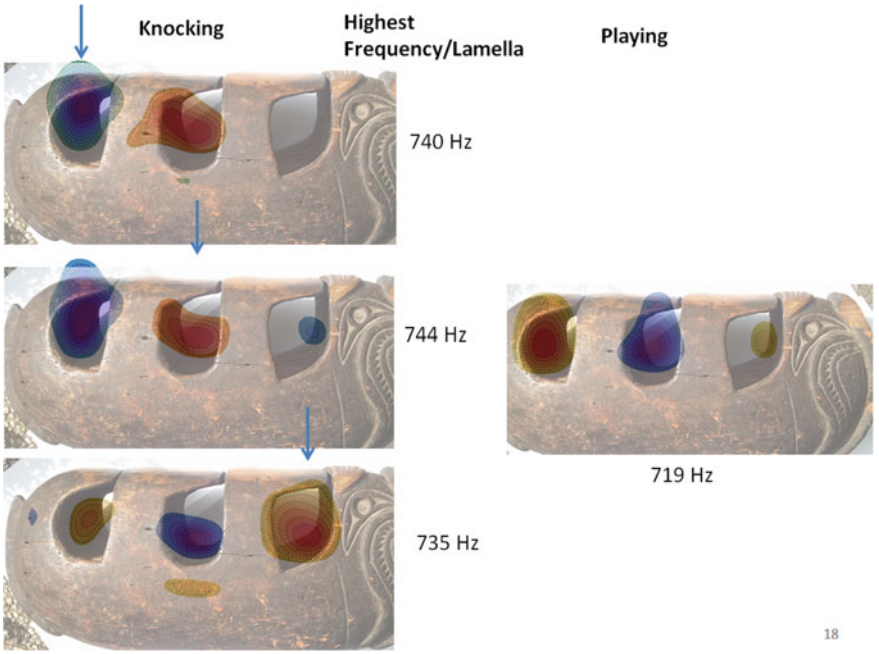


Fig. 4 Comparison between knocking on different plates and playing the lowest plate for the lowest partial in the respective spectra. The dependency between the driving point and the radiation pattern clearly appears. The resonance needs all three chambers where the phase-antiphase relation changes regularly between the chambers

6 Physical Modeling of Burmese *Pat Wain* Drum Tuning

As an example of the use of Physical Modeling the *pat wain* drum circle of the Bama *hsain wain* orchestra is analyzed. Only when investigating the drum heads in such a way it is possible to understand both, the tuning of the instrument as well as the question how a percussion instrument like the drum can be used as a melody instrument at all.

The *pat wain* is a drum circle unique in the world, consisting of 20 or 21 pitched drums. The instrument is the leader in a Bama *hsain wain* orchestra used for entertainment, puppet theater, or weddings [67, 68]. Basically the orchestra is a drum-and-shawm ensemble found all over Southeast Asia, India, Sri Lanka or China which consists of a *zurna*- or oboe-like instrument, here called *hne* and drums, here the six drum set *chauklon pat* and the big drum *sakhun*. Additionally there are gong chimes, a circular *ci wain* and/or *maun sain*, a chime set resting in a box. A clapper and a cymbal are used for additional rhythmic coloration.

This orchestra type was introduced by the Janizary military band of the Osman empire and fits well the needs of an outdoor band, it is loud and provides melody and rhythm. The *hsain wain* ensemble of Myanmar is a native orchestra of the Bama, the

main ethnic group in multi-cultural Myanmar which officially has over 130 ethnic groups. Still it is also played outside the Bama mainland of Yangon or Mandalay. The recordings used in the present paper are taken from Kay Zaw, a *pat wain* player in Myitkyina, the capital of nothern Myanmar Kachin state.

The *pat wain* is tuned by a paste consisting of rice and ashes. In the region of Myitkyina two kinds of rice are planted, normal wet rice and a mountain dry rice. Cooking so-called sticky rice, a sweet, is common practice and this kind of rice is especially sticky, which means it has a high amount of viscoelasticity. The tuning process is performed by adding more and more paste to the center of the drum head, covering about half of the drum. The more paste is applied the lower is the sound, as additional mass without a considerable change of the membrane tension must lead to a lowering of the drums eigenfrequencies. Then after applying some paste it is redistributed over the drum head, while the tuner is constantly checking the timbre and comparing it to the other drums in the set. The checking of the timbre of the drum therefore also is a crucial part of the tuning process. So the tuning has two aims, to pitch the instrument in terms of its fundamental and to tune its overtone spectrum by redistributing the paste.

Adding additional paste to a drum head is a common practice for drums. Indian drums, especially *tablas* are very well studied in terms of the additional paste applied to their drum head [59, 69] called *sihai* which is placed concentric for the *dayan* and off-centric for the *bayan* drum. Varying the width, position, smoothness and strength of the *sihai* it was found that the harmonic overtone relations of the drum modes change and can meet values very close to harmonic ratios.

Tablas have a clear pitch, still they are not played as melody instruments. The *pat wain* on the other hand needs to be tuned to pitches over about three octaves. The present study analyzes the drums of a *pat wain* drum set and compares the results to a Finite-Difference physical model of a single drum, as well as a coupling of two drum heads. Experimentally it was found that a double-headed drum, like the *pat wain* shows a coupling of the drum heads for lower modes, while the higher modes are more or less decoupled [59, 70]. Normally the upper drum head is keeping its frequency of the lower modes while the lower drum head is forced into the motion and frequency of the upper one. As we will see with the *pat wain* there are only two or three lower modes left in the sound and therefore all modes are expected to be coupled between the two drum heads.

Several other studies deal with drums. The vibration of the Karen bronze or ‘frog’ drum has been studied experimentally [71], finding complex modes up to 3 kHz. The influence of non-uniform tension of a drum head was studied using laser-interferometry measurements, where degenerated modes have been found [72] mostly leading to musical beats. The eigenmodes of the drum vessel were analyzed using Finite-Element methods [73]. The coupling between the drum head and the wooden shell was investigated using Finite-Elements for a bass drum [74].

As the recordings were performed during a field trip to Nothern Myanmar in 2014 the drums could not be investigated in the lab. Still Kay Zaw, the musician was tuning the drums for the recording for about an hour very carefully in this hut next to his house and therefore original conditions were present in terms of humidity,

nearly a hundred percent, and temperature, which was about 35 °C. Suzuki found a strong dependency of the drum modes in terms of temperature and humidity [75] for a Japanese drum and therefore original conditions need to be present in the study.

Additionally the internal damping of the drum heads is considered. The *pat wain* drums are made of leather of deers or the water buffalo which has strong internal damping [76]. Leather consists of collagen which is a very stiff protein and therefore leather is basically considered a crystal. Still the layering of leather allows gaps which are filled by water, as well as by molecules which are introduced by the process of tanning. Here water molecules enter in two ways, either as larger portions between collagen fibers or as single molecules between or even within collagen molecules. The former is responsible for leather to freeze to a hard plate around zero degree celsius. The other water molecules are melting slowly with temperature adding a lot to the strong viscoelastic properties of leather. Additional components of this viscoelastic internal damping are sudden changes in the molecular geometries when stress is applied. All these processes lead to a phase shift between stress and strain which again leads to an internal energy loss of the vibration. Internal damping can be very strong and therefore contributes considerably to the timbre of the drum sound. If and to which extend this internal damping also contributes to the change of frequencies is modeled using the viscoelastic Finite-Difference model proposed in this paper.

Physical models of the drum have been performed for Indian drum heads [69], the snare drum [77], or the bass drum [74]. Basic models of viscoelastic damping have mainly been discussed with Finite-Element Methods [78]. Here the Maxwell and the Kelvin-Voigt model for relaxation and creep respectively are normally used as time integration models, where combinations of both are able to build arbitrarily complex damping behaviour.

The additional paste adds considerable damping to the system. Performing a viscoelastic damping model as part of a Finite-Difference model the frequency change of partials due to heavy damping is estimated.

6.1 Finite-Difference Model of the Drum Head

The drum membrane is modeled as a Finite-Difference Time Domain (FDTD) model implementing the equation for a membrane with tension T, area density $\mu(x, y) = m(x, y)/A$ and damping constant D with displacement u, like

$$\frac{T}{\mu(x, y)} \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right) = \frac{\partial^2 u}{\partial t^2} + D \frac{\partial u}{\partial t} . \tag{1}$$

The area density depends on the mass m(x, y) which has a spatial distribution according to the additional paste. In the present study a doubling of the mass compared to the case without paste is assumed, according to

$$\mu(x, y) = m(x, y)/A = \begin{cases} \mu_0 \text{ for } \sqrt{(x - x_0)^2 + (y - y_0)^2} > 0.1\ n\ r \\ 2\ \mu_0 \text{ for } \sqrt{(x - x_0)^2 + (y - y_0)^2} < 0.1\ n\ r\ , \end{cases} \tag{2}$$

with n = 0, 1, 2, … 10 and membrane center position (x_0, y_0). So a circular paste area is modeled which covers the membrane from zero radius (no coverage) up to full radius (full coverage) in ten equidistant radius steps.

When discussing modes, a model with a single membrane was used. Still to consider the influence of the back membrane and the inclosed air on the fundamental frequency amplitudes, the back membrane and the air are included in the model. The back membrane is modeled like that of the front membrane with the exception that when adding viscoelastic damping, as discussed above, this damping was omitted for the back membrane. This choice follows the result discussed below, that viscoelastic damping does not change low frequencies considerably.

The air was not modeled as a discrete model but it was assumed that the front and back membrane interact with a delay caused by the traveling of waves through the height of the drum. The coupling between the membranes and the air was modeled according to the Euler equation connecting membrane velocity and air pressure.

6.2 Viscoelasticity Model

The differential equation of the membrane discussed above includes a damping term. The reason for this damping is not clearly defined. Principally there is external damping caused by energy loss due to radiation and internal damping caused by energy loss within the structure. The reasons for the internal energy loss are not perfectly clear. There are thermodynamic losses of several kinds, viscoelastic losses due to shearing, atomistic and quantum mechanical considerations of molecular restructuring and several others [79]. Experimental data of viscoelastic losses in leather show an increase of damping with higher temperatures above the glass transition temperature of leather [76] which are additionally frequency dependent. Higher water content of leather also leads to higher internal damping. The reasons for such damping behaviour in leather are mainly thought to be a reconfiguration of collagen fibers within and between the fibers and with the water molecules building part of the collagen structure.

Discussing internal damping is beyond the scope of the present paper. Still all explanations of internal damping end up in a simple model of a complex and frequency dependent Young's modulus E(s) with the complex frequency

$$s = \alpha + \imath\omega. \tag{3}$$

Then the stress-strain relation in the frequency domain becomes

$$\sigma(s) = E(s)\ \epsilon(s), \tag{4}$$

with stress σ and strain ϵ. To implement this in a model, this equation can be transformed into the time-domain using a Laplace transform like

$$\sigma(t) = \int_0^\infty \epsilon(t-\tau)\, h(\tau)\, d\tau. \tag{5}$$

Here $h(\tau)$ is a function representing the time domain of the complex Young's modulus E(s). So the present stress is the result of all previous strains weighted with $h(\tau)$. Therefore each spectral component of a sound is damped with its own damping parameter $\alpha(\omega)$ and therefore has a time series like

$$u_\omega(t) = A(\omega)\, e^{\alpha\, t}\, e^{\imath\, \omega\, t}. \tag{6}$$

This means that there is a temporal delay between stress and strain which can be expressed as an angle δ, the phase relation between stress and strain for a single frequency. δ is often measured as this phase relation and in the literature often written as

$$\tan\delta = E_I / E_R, \tag{7}$$

the relation between the imaginary and real parts of the complex Young's modulus.

The model was included in the study as the paste, consisting of rice and ashes is expected to have a strong viscoelasticity, additional to that of the leather drum head. So then it can be discussed if a strong viscoelastic damping will also change the frequency of the drum head and therefore help tuning the instrument, or if the influence is too small to be of relevance.

The stress-strain relation is also the definition of the Young's modulus. In the case of a membrane we do not have a Young's modulus, so we need to transfer the idea. Strain is dimensionless and refers to the potential energy of the system caused by displacement differences. The stress is weighted force applied to the structure in order to obtain the strain. In the dynamical case this force can have different parts, the acceleration of the system, damping or external forces. The unit of the Young's modulus is that of the stress, force over area, as the strain is dimensionless.

So the stress-strain relation is a force balance, according to Newton's idea of mechanical systems in which all interactions can be written as a sum of forces. Therefore it is straightforward to replace the strain with the spatial differentiation of the membrane and the Young's modulus by force over area density. As the area density is hardly complex, clearly the tension is the parameter we can refer to as complex and frequency dependent. Then the viscoelastic membrane equation reads

$$\int_{\tau=0}^{T} h(\tau)\frac{T}{\mu(x,y)}\left(\frac{\partial^2 u(x,y,t-\tau)}{\partial x^2}+\frac{\partial^2 u(x,y,t-\tau)}{\partial y^2}\right) = \frac{\partial^2 u(x,y,t)}{\partial t^2}+D\,\frac{\partial u(x,y,t)}{\partial t}.$$

This equation was implemented in the viscoelastic model.

The calculations were performed on a Graphics Processing Unit (GPU) with massive parallel computation. Still the model is not real-time and—depending on the GPU used, calculating one second of sound with a sample rate of 96 kHz takes between 5–10 s.

6.3 Partials of Drum Sounds

To make a drum a pitched instrument at least one strong partial should be in simple harmonic relation to the fundamental, like 2:1 or 3:1. It should also not contain too many inharmonic partials as again a clear pitch would be hard to perceive. Then it would be an advantage to have the drum sound for a time span which gives a listener the chance to really perceive a pitch but not too long to disturb following notes, as otherwise a played drum would need to be damped before playing a next one. An example of an instrument where such damping is necessary is the Balinese *gender dasa* as used for *wayang kulit* puppet theater. This instrument consists of bronze plate with low internal damping which sound very long. So before striking a note the previous plate needs to be damped not to blur the sounds and pitches which is achieved by the fingers grabbing the plate.

So the acoustics of the *pat wain* need to meet these constraints for the instrument to be usable as a pitched instrument at all. When measuring the partials of the sounds is appears that at least in the middle range of the scale the first two overtones come very close to a ratio of 2:1 and 3:1, where the 3:1 is met even better. Indeed, when investigating the decay of these partials, the third partial is mostly much stronger in amplitude than the second. The other partials are only present within the initial transient phase of about 50 ms or less and therefore do not disturb the pitch impression. Also the fundamental is vibrating much longer than the overtones which ends in a nearly sinusoidal sound and again enhances pitch perception.

Figure 5 shows the measured frequencies of the first three partials for the drums. The two additional smooth curves shows the resonance frequencies of the inclosed air volume of the drums estimated as a half wavelength (light blue) and a whole wavelength (green) fitting into the body of the drums. The precise drum heights are not known, still they range from 13 cm for the highest and 41 cm for the lowest drum. It appears that these resonances are around the second and third partial and are therefore able to boost these partials in terms of making their decay longer.

Figure 6 shows the relations between the first and second partial (light blue) and the first and third partial (green). In a middle range of the drums these relations come

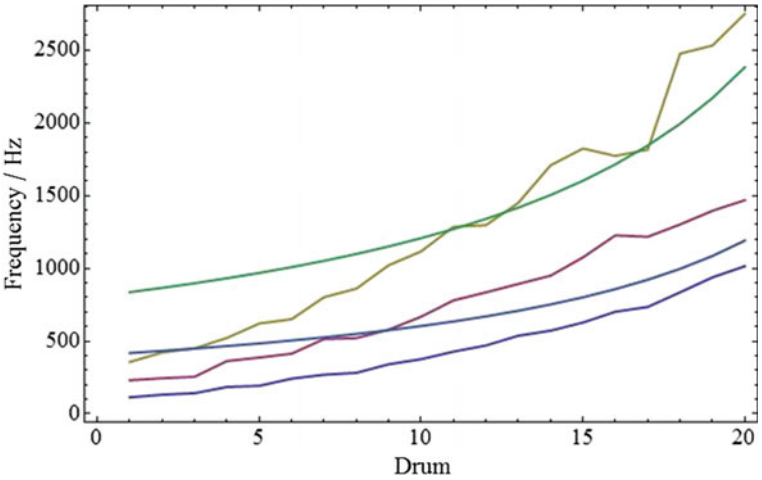


Fig. 5 Frequencies of the first three *pat wain* drum partials. The light blue and the green curve show the estimated resonance frequencies of the air volume inside the drum when taking the drum height as half wavelength (light blue) or whole wavelength (green). It appears that the air volume supports both frequencies round about

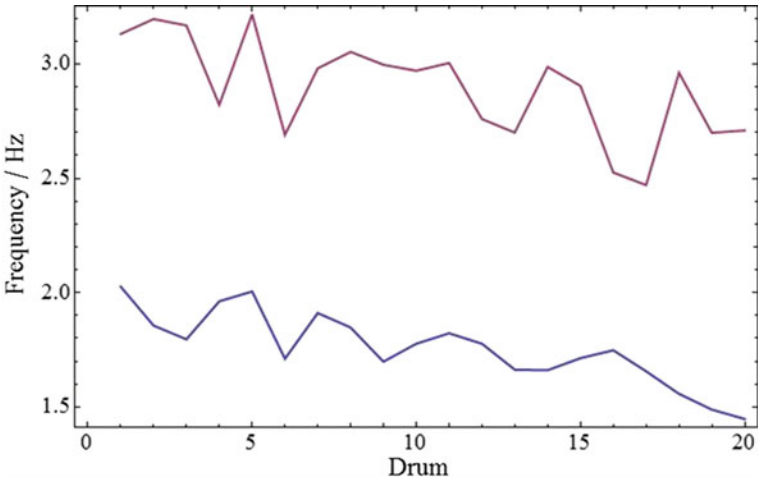


Fig. 6 Relations between the first and second (blue) and the first and third (red) partials taken from Fig. 5. Within a middle range the relations are close to 2:1 and even better 3:1

close to 2:1 and, even better 3:1. As these are the partials which are also boosted by the air volume of the drums they are both, quite strong in terms of amplitude and within about harmonic relations to the fundamental.

This seems to be the basic principle behind these *pat wain* drums, making them suitable for pitched playing, to have at least two nearly harmonic and strong harmonic partials.

6.4 Modes of Membrane for Different Paste Radius

To decide if the amount of paste applied to the drumhead is of importance in terms of the relation of the first three partials, in the model the amount of paste on the membrane has been changed, changing $\mu(x, y)$ in 11 cases. As applying paste lowers the frequencies, sorting the results according to frequency means that the first case is when the past fully covers the membrane. Then in ten steps the amount of paste is reduced until in the last case no paste is applied. As the paste is applied in a circle centered in the middle of the membrane, the ten cases reduce the area of paste coverage by 1/10th of the membrane radius. So then the 6th case from the highest frequency with no paste means the paste covers 0.5 of the membrane.

Figures 7 and 8 show the modes of the lowest partials for all these cases. The lowest frequency with full paste coverage is on the very left, the highest frequency with no coverage is on the right. The 6th case from the right is 0.5 coverage.

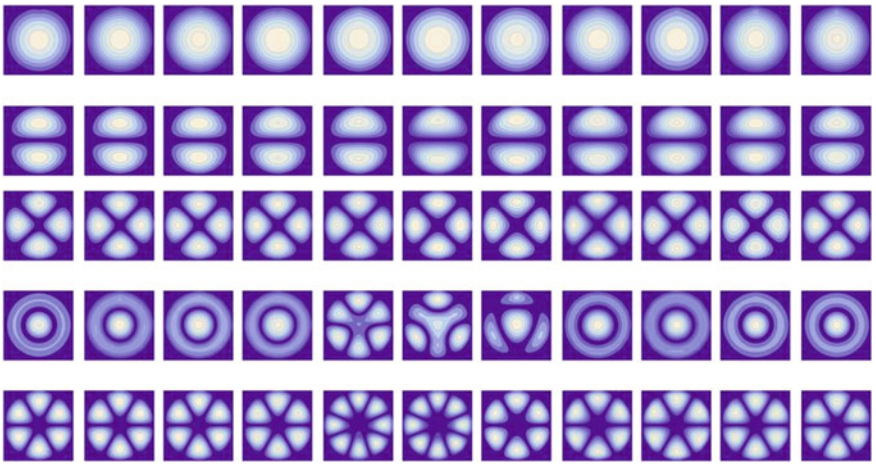


Fig. 7 Modes of the first five frequencies of the membrane with 11 cases of pate applied to the drum head. The very right case has no paste and each adjacent case on the left has increasing radius of $n/11$ as circle centered at the membrane center. So the sixth case from the right covers 0.5 of the membrane, the very left case covers all of the membrane

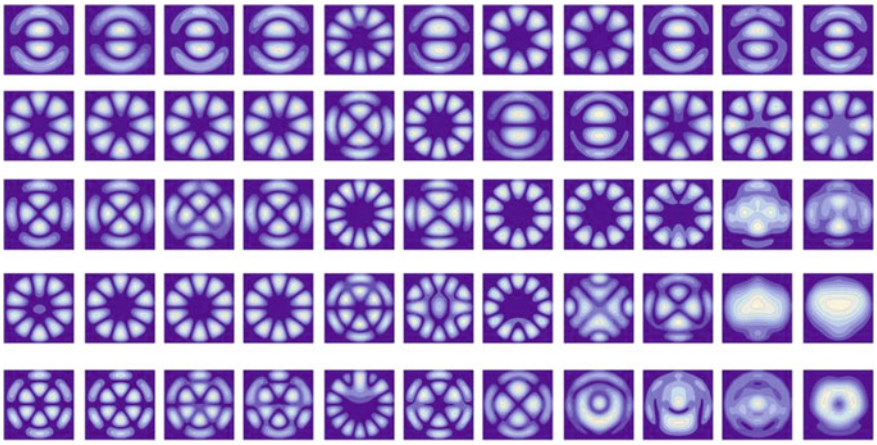


Fig. 8 Same as Fig. 7, now for partials six to ten

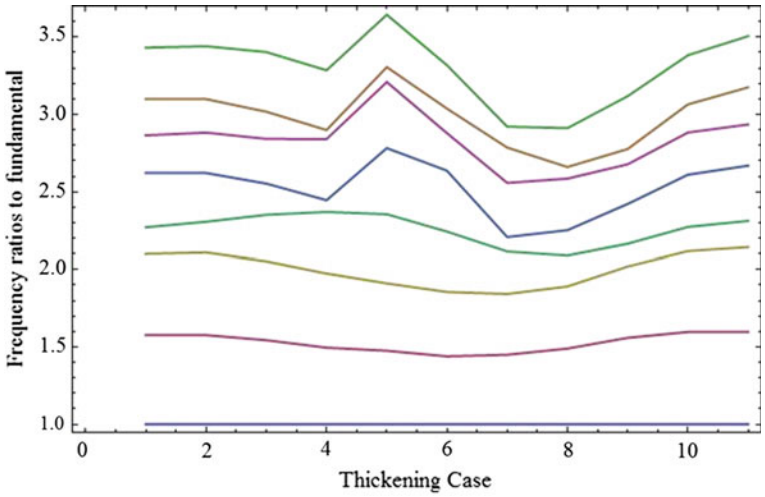


Fig. 9 Relations of modes with respect to lowest frequency of the modes shown in Figs. 7 and 8 for the ten cases of paste. Again the very right case is no paste, the very left is full paste and case six from the right covers 0.5 of the membranes radius with paste

It appears that all modes change through the different steps, still the lowest three modes do not change their basic mode shape but only show slight changes in the displacement distribution. Starting from the third mode, in the middle pate radius range, the range which the musicians actually use, the basic modes shapes change. There even is a flip of modes, as is the case with the fourth and fifth cases from the right for modes six and seven. The ordering of modes in all cases is according to their frequencies. In some cases, modes no longer appear, as e.g. the (1, 1) mode of the

seventh case from right which is the sixth mode of the cases on the left. The vanishing of clear mode shapes to the very right lower corner is due to reduced amplitudes of these cases, where the sound is so short that these modes do not have a chance to build up completely.

Homogeneous, isotropic, circular membranes have occasional frequency relations of about 2:1 of mode (2, 0) of 2.136:1 and 3:1 for modes (1, 1) with 2.918:1 and (4, 0) with 3.156:1. Still when applying paste these relations change, as do the modes themselves. Figure 9 shows the relations of the modes of the previous figures to their fundamental. The mode close to 2:1 is moving its relation around this 2:1 which means that applying paste is lowering the relation. With 0.5 coverage the relation is a bit below 2:1. With the modes near 3:1 there is a considerable frequency change due to the mode change. The frequency relations lower, raise again and suddenly flip to a lower relation when applying more and more paste. Indeed, in the sixth case with 0.5 coverage the frequency relations are both very close to 3:1.

So when covering the membrane half with paste, as applied by musicians, the third partial, which is the loudest one in the overtone spectrum is tuned around 3:1. Still for the 2:1 relation a coverage of one third or two thirds would be preferable.

Pat wain players tune the drum head in two ways, first they apply a bit of paste lowering the sound and then they distribute the paste over the membrane and keep testing the sound. We have seen that different distributions of the paste will change the overtone relations. So it might be that after hitting the right fundamental pitch the players make a fine-tuning of the second and third partial to reach a sound which has a strong pitch character. Unfortunately, as in many cases with musicians, the players are not perfectly aware of what they do and cannot tell and rationalize their tuning process. So although theoretically such a tuning is possible and desired, the question if it is really applied by the musicians is not finally answered here. A detailed investigation of different tuning stages would need to be done which is hard to do in a real fieldwork.

So only when including Physical Modeling to a Phonogram Archive it is possible to get insight into tuning details as well as the possibility to use the *pat wain* as a melody instrument at all.

7 Conclusion

A Computational Music and Sound Archive does not only fit the needs of sorting and analyzing music automatically in times of Big Data and digital accessibility of music. It also is a way to understand music and culture in terms relating cultures and ethnic groups rather than stress the dissimilarities between them. It is also objective in a way to omit cultural bias and view. Therefore it suggests a view on music as a complex system rather than sorting it in genres or styles. Such a Physical Culture Theory is therefore both able to cope with the complexity of todays reality as well as suggesting a new and fresh look on music in the world.

References

1. Schneider A (2018) Systematic musicology: a historical interdisciplinary perspective. In: Bader R (ed) Springer handbook of systematic musicology. Springer, Heidelberg, Berlin, pp 1–24
2. Schneider A (2006) Comparative and systematic musicology in relation to ethnomusicology: a historical and methodological survey. Ethnomusicology 50(2):236–258
3. Heins E, den Otter, E, van Lamsweerde F (1994) Jaap Kunst: traditional music and its interaction with the West, KIT Publishers
4. Hornbostel EMV, Sachs C (1914) Systematik der Musikinstrumente. Ein Versuch. Zeitschrift für Ethnologie 46:4–5, 553–590
5. Sachs C (1913) Real-Lexikon der Musikinstrumente: zugleich ein Polyglossar fr das gesamte Instrumentengebiet. Bard, Berlin
6. Andre R (1998) Elektronische Klnge und musikalische Entdeckungen [Electronic Sounds and musical discoveries]. Reclam, Ditzingen
7. Schneider A (2001) Sound, pitch, and scale: from tone measurements to sonological analysis in ethnomusicology. Ethnomusicology 45(3):489–519
8. Lomax A, Berkowitz N (1972) The evolutionary taxonomy of culture. Science 177(4045):228–239
9. Lomax A (1976) Cantometrics: an approach to the anthropology of music. University of California Extension Media Center, Berkeley
10. Bader R (2013) Nonlinearities and synchronization in musical acoustics and music psychology. In: Springer series current research in systematic musicology, vol 2. Springer, Heidelberg
11. Kostek B (2005) Perception-based data processing in acoustics. In: Applications to music information retrieval and psychophysiology of hearing, Springer
12. Leman M, Carreras F (1997) Schema and gestalt: testing the hypothesis of psychoneural isomorphism by computer simulation. In: Leman M (ed) Music, gestalt, and computing studies in cognitive and systematic musicology. Springer, Berlin, pp 144–168
13. Toiviainen P (ed) (2009) Musical similarity [Special issue]. Musicae Scientiae, vol 13(1 suppl)
14. Toiviainen P, Eerola T (2001) A method for comparative analysis of folk music based on musical feature extraction and neural networks. In: Lappalainen H (ed) Proceedings of the VII international symposium of systematic and comparative musicology and the III international conference on cognitive musicology. University of Jyvskyl, Jyvskyl, pp 41–45
15. Downie JS (2003) Music information retrieval. Annual Rev Inf Sci Technol 37:295–340. http://music-ir.org/downie_mir_arist37.pdf
16. Fingerhut M (2004) Music information retrieval, or how to search for (and maybe find) music and do away with incipits. In: Proceedings IAML IASA joint congress, music and multimedia, Oslo, Aug 2004. http://www.ismir.net/admin/ismir-booklet.pdf
17. Klapuri A, Davy M (2006) Signal processing methods for music transcription. In: Klapuri A, Davy M (ed) Signal processing. Springer, New York Inc
18. Alexandraki CH, Bader R (2013) Real-time concatanative synthesis for networked musical interactions. J Acoust Soc Am 133:3367
19. Emmanouil B, Simon D (2013) Multiple-instrument polyphonic music transcription using a temporally constrained shift-invariant model. J Acoust Soc Am 133(3):1727–1741
20. Rohrmeier M, Pearce M (2018) Musical syntax I: theoretical perspectives. In: Bader R (ed) Springer handbook of systematic musicology. Springer, Berlin, Heidelberg, pp 473–486
21. Pearce M, Rohrmeier M (2018) Musical syntax II: empirical perspectives. In: Bader R (ed) Springer handbook of systematic musicology. Springer, Berlin, Heidelberg, pp 487–505
22. Rohrmeier MA, Cross I (2014) Linking implicit and statistical learning. Modelling unsupervised online-learning of artificial grammars. Conscious Cognit Int J 27:155–167
23. Zucchini W, MacDonald IL (2009) Hidden-Markov models for time series. An introduction using R, Chapman & Hall
24. Bader R Computational mechanics of the classical guitar. Springer, Oct 2005
25. Gary S (2018) Delay-lines and digital waveguides. In: Bader R (ed) Springer handbook of systematic musicology. Springer, Berlin, Heidelberg, pp 259–272

26. Fischer J (2017) Numerical simulations of the turbulent flow and the sound field of the Turkish Ney end-blown flute. J Acoust Soc Am 141:39–60
27. Pfeifle F (2018) Real-time signal processing on field programmable gate array hardware. In: Bader R (ed) Springer handbook of systematic musicology. Springer, Berlin, Heidelberg, pp 385–417
28. Florian P, Bader, R (2015) Real-time finite-difference method physical modeling of musical instruments using field-programmable gate array hardware. J Audio Eng Soc 63(12):1001–1016
29. Bader R (2014) Microphone array. In: Rossing T (ed) Springer handbook of acoustics, pp 1179–1207
30. Thomas M (2018) Measurement techniques. In: Bader R (ed) Springer handbook of systematic musicology. Springer, Berlin, Heidelberg, pp 81–103
31. George B (2003) Modal analysis of the violin octet. J Acoust Soc Am 113(4):2105–2113
32. Plath N (2013) High-speed camera displacement measurement (HCDM) technique of string vibrations. In: Proceedings of the Stockholm music acoustics conference, pp 188–192
33. Copeland P (2008) Manual of analogue sound restoration techniques. The British Library, London
34. Elschek O et al (2001) Digitizing world music. Digitalisierung von Weltmusik. Special issue, Systematische Musikwissenschaft/Systematic Musicology, vol VII, no 3
35. Proutskova P (2007) Musical memory of the world data infrastructure in ethnomusicological archives. In: Proceedings of the 8th international conference on music information retrieval (ISMIR), Vienna, Austria
36. The National Recording Preservation Board of the Library of Congress (2010) The state of recorded sound preservation in the United States: a national legacy at risk in the digital age. Washington, D.C. Council on Library and Information Resources and The Library of Congress
37. The National Recording Preservation Board of the Library of Congress (2012) The library of congress national recording preservation plan. Washington, D.C. Council on Library and Information Resources and The Library of Congress
38. Kolinski M (1978) The structure of music: diversification versus constraint. Ethnomusicology 22(2):229–244
39. Gmez E, Herrera P, Gmez-Martin F (2013) Compuational ethnomusicology: perspectives and challenges. J New Music Res 42(2):111–112
40. Pisczcalski M, Geller B (1977) Automatic music transcription. Comput Music J 1(4):24–31
41. Klapuri A (2004) Signal processing methods for the automatic transcription of music. PhD dissertation, Tampere University of Technology
42. Goto M (2001) An audio-based real-time beat tracking system for music with or without drum-sounds. J New Music Res 30(2):159–171
43. Klapuri A (1999) Sound onset detection by applying psychoacoustic knowledge. In: 1999 IEEE international conference on acoustics speech and signal processing proceedings, ICASSP99 Cat No99CH36258 6, pp 3089–3092
44. Dessein A, Cont A, Lemaitre G (2010) Real-time polyphonic music transcription with non-negative matrix factorization and beta-divergence. In: Proceedings of the 11th international society for music information retrieval conference (ISMIR), pp 489–494
45. Foote J (2000) Automatic audio segmentation using a measure of audio novelty. In: Proceedings ICME '00
46. Foote J, Cooper M (2003) Media segmentation using self-similarity decomposition. In: Proceedings of SPIE storage and retrieval for multimedia databases, vol 5021
47. Six J, Cornelis O, Leman M (2013) Tarsos, a modular platform for precise pitch analysis of western and non-western music. J New Music Res 42(2):113–129
48. Dixon S (2006) Onset detection revisited. In: Proceedings of the international conference on digital audio effects DAFx06, pp 133–137
49. Goto M (2001) A predominant-f0 estimation method for real-world musical audio signals: MAP estimation for incorporating prior knowledge about f0s and tone models. In: Proceedings workshop on consistent and reliable acoustic cues for sound, pp 1–4

50. Bader R (2011) Buddhism, animism, and entertainment in Cambodian melismatic chanting SMOT. In: Schneider A, von Ruschkowski A (eds) Hamburg yearbook of musicology, vol 28
51. Chai W, Vercoe B (2003) Structural analysis of musical signals for indexing and thumbnailing. In: Proceedings ACM/IEEE joint conference on digital libraries
52. Gibiat V, Castellengo M (2000) Period doubling occurences in wind instruments musical performance. Acustica 86:746–754
53. Leman M (1995) Music and schema theory. Springer, Berlin
54. Rabiner LR, Juang BH (1993) Fundamentals of speech recognition, Prentice Hall Signal Processing Series
55. Aucouturier J-J, Sandler M (2001) Segmentation of musical signals using hidden Markov models. In: Proceedings of the audio engineering society 110th convention
56. Bader R (2018) Springer handbook of systematic musicology. Springer, Berlin, Heidelberg
57. Fletcher N, Rossing THD (2000) Physics of musical instruments. Springer, Heidelberg
58. Rossing THD (2010) Science of stringed instruments. Springer, New York, Heidelberg
59. Rossing Thomas D (2000) Science of percussion instruments. World Scientific, Singapore
60. Bader R, Hansen U (2008) Acoustical analysis and modeling of musical instruments using modern signal processing methods. In: Havelock D, Vorlnder M, Kuwano S (eds) Handbook of signal processing in acoustics. Springer, pp 219–247
61. Sadie S (1984) The new grove dictionary of musical instruments, vol 2. Macmillan Press Limited
62. Woodhouse J, Alluzzo PM (2004) The bowed string as we know it today. Acta Acustica United with Acustica 90:579–589
63. Woodhouse J, Schumacher RT (1995) The transient behaviour of models of bowed-string motion. Chaos 5:509–523
64. Heintze D (2011) Lounúet. Notizen zum neuirländischen Reibidiophon. [Lounúet. Notes about the friction idiophone from New Ireland.]. In: Deterts D, Heintze D, Seybold S (eds) Musik—Ethnologie—Museum. Überseemuesum Bremen, Schünemann, Freundesgabe für Andreas Lüderwaldt. Jahrbuch XVII, pp 69–104
65. Messner F (1998) Friction blocks of New Ireland. In: Kaeppler AL, Love JW (eds) Garland encyclopedia of world music. Australia and the Pacific Islands, vol 9. Routledge, London, pp 380–382
66. Messner GF (1980) Das Reibholz von New Ireland. Manu Taga Kul Kas...' (Der "Vogel" singt noch...) [The friction wood from New Ireland. Manu Taga Kul Kas...' (The 'bird' still sings...)]. Studien zur Musikwissenschaft, Band 31, pp 221–312
67. Wolf D (2010) Bamarische Musik. Yodaya Lieder im kulturell-historischen Kontext Myanmars [Bama music. Yodaya songs in the cultural-historical context of Myanmar.] regiospectra-Verlag Berlin
68. Zaw UK (1981) Burmese culture, general and particular. Sarpay Beikman, Printing and Publishing Corporation, Ministry of Information, Rangoon
69. Sathej G, Adhikari R (2009) The Eigen spectra of Indian musical drums. J Acoust Soc Am 126(2):831–838
70. Worland R (2011) Demonstration of coupled membrane modes on a musical drum. J Acoust Soc Am 130:2397
71. Nickerson LM, Rossing ThD (1999) Acoustics of the Karen bronze drums. J Acoust Soc Am 106:2254
72. Worland R (2010) Normal modes of a musical drum head under non-uniform tension. J Acoust Soc Am 127(1):525–533
73. Hwang Y-F, Suzuki H (2016) A finite-element analysis on the free vibration of Japanese drum wood barrels under material property uncertainty. Acoust Sci Tech 37(3):115–122
74. Bader R (2006) Finite-element calculation of a bass drum. J Acoust Soc Am 119:3290
75. Suzuki H, Miamoto Y (2012) Resonance frequency changes of Japanese drum (nagado daiko) diaphragms due to temperature, humidity, and aging. Acoust Sci Tech 33(4):277–278
76. Jeyapalina S (2004) Studies of the hydro-thermal and viscoelastic properties of leather. Univ of Leichester, PhD

77. Bilbao S (2012) Time-domain simulation and sound synthesis for the snare drum. J Acoust Soc Am 131(1):914–925
78. Wriggers P (2001) Nichtlineare finite-element methoden. [Nonlinear Finite-Element Methods], Springer
79. Pierce A (2010) Intrinsic damping, relaxation processes, and internal friction in vibrating systems. POMA 9:1–16

Part II
Fieldworks and Archives

"The *Lanang* Is the Bus Driver": Intersections of Ethnography and Music Analysis in a Study of Balinese *Arja* Drumming

Leslie Tilley

Abstract This chapter presents an ethnographically informed analysis of the improvised Balinese drumming practice *kendang arja*, blending a personal and sometimes informal ethnographic writing style with close musical analysis. In this paired drumming tradition, two musicians, without formal music theory to guide them, somehow create the tightly interlocking rhythms so characteristic of Balinese music through continuous and simultaneous improvisation. No surprise these are among Bali's most respected musicians. By examining the improvisations of various master drummers via the lens of their informal oral music theory, this chapter seeks to make explicit several relatively implicit guidelines and techniques for *kendang arja* improvising. It thus offers insight into the ways these musicians create mutually compatible patterns in the course of performance, while simultaneously demonstrating the value of ethnographic fieldwork to the analysis of improvised musics.

1 Setting the Scene

"I want you to perform with me at the temple tomorrow night," my teacher Pak Tama casually informs me as I kick off my flip flops and join him on the porch where we have our daily drum lessons. To me, these are the scariest twelve words he could have uttered! The music I've been studying with him, *kendang arja,* demands that two drummers simultaneously improvise intricate interlocking patterns as the accompaniment to an epic sung dance-drama. Longtime partners develop their own personal styles, learning to read subtle rhythmic and physical cues, and each can respond to the other's musical offerings in real time with astonishing ease and panache. As in the

Electronic supplementary material The online version of this chapter (https://doi.org/10.1007/978-3-030-02695-0_2) contains supplementary material, which is available to authorized users.

L. Tilley (✉)
Massachusetts Institute of Technology, Cambridge, MA, USA
e-mail: tilley@mit.edu

R. Bader (ed.), *Computational Phonogram Archiving*, Current Research in Systematic Musicology 5, https://doi.org/10.1007/978-3-030-02695-0_2

Fig. 1 Longtime partners play with ease. I Dewa Nyoman Sura (Pak Dewa) of Pengosekan (left) and I Cokorda Alit Hendrawan (Cok Alit) of Peliatan (right) (Photo by Chelsea Edwardson, 2011. Used with permission)

photo in Fig. 1, good *arja* drummers make the seemingly impossible look effortless.

Younger drummers raptly watch these older masters, hoping to catch just a snippet of a newly improvised pattern or understand the spontaneous construction of a perfect moment of interlocking. Performing this high-stakes drumming at an important religious ceremony was not what I had in mind when I woke up this morning! So far I have learned just two 8-beat patterns as a basis for *arja* improvising. And I don't yet understand how to create new idiomatic patterns on my own. What's more, I've been studying the higher *lanang* drum, the leader of the ensemble, and this is what Pak Tama has asked me to play in the temple.

"I don't think I'm ready yet," I reply self-consciously. How can a leader play just two patterns over and over and call it "improvisation"? But Pak Tama is unmoved by my protestations. "Don't worry. You'll do fine," he smiles. "Remember, the *lanang* is the bus driver." Then he picks up his drum and indicates for me to do likewise, the matter now apparently settled.

Though this somewhat cryptic exchange offers me little comfort at the time, it—and others like it—will later fundamentally shape my analysis of the practice.

2 The Research Question

Balinese music is famous for its interlocking techniques, and the paired conical drumming traditions are no exception. Strokes on the higher-pitched *kendang lanang* intertwine with patterns of like strokes on the lower-pitched *kendang wadon* to create complex composite patterns. Almost invariably, these patterns are exactingly composed. In the drumming for the Balinese dance-drama *arja*, however, interlocking is created through improvisation (see Video 1). How two simultaneously improvising drummers are able to weave their patterns so fluidly around one another at such high speeds is an analytical question that has only begun to be investigated (see [12]). *Arja* drumming has not been theorized in a formal institutional context; the social construction of its materials is ad hoc, and varies by teacher and place. Yet the drummers' consistent ability to interlock with their partners in the course of performance implies a set of guidelines being followed, whether consciously or not. What insights can an ethnographic approach to *kendang arja* offer a music analyst? How can conversations, lessons, and performances with expert musicians elucidate their musical processes and techniques? And in what ways can the oral music theory we find in these contexts, informal and often ambiguous though it may be, guide and shape a close musical analysis of their improvisations?

2.1 The Importance of Fieldwork to Ethnomusicological Analysis

Most scholars of music begin to learn analysis techniques in first-year music theory classes, where simple rules of structure, harmony, and voice-leading illuminate basic compositional procedures through the close examination of musical scores. As beginners, our analyses are inextricably linked to notation. Yet while a physical score can be an indispensible tool for a music analyst, enabling slow, meticulous scrutiny of musical features, the idea that analysis is meant to elucidate a musical "work" can be problematic when researching aural music cultures. Recent musicological studies are now "re-framing […] the traditional roles of objects and practitioners in the work of music, and in the work of musicology" [17: 3]. We have come to see music as a *process* as much as a *work* (see [6, 7]). Despite such nuancing, however, both transcription and analysis have been the topic of much debate among ethnomusicologists since the mid-twentieth century, when concerns about the power dynamics of representing one culture's music through the approaches of another were brought to the fore (see [9, 14, 18]). Even if we are able to transcribe the music we are researching in appropriate, relatively culturally faithful ways, we still must interpret the information we find there, determining which details are most meaningful. We can more precisely train our analytical lenses by keeping our teachers and collaborators within the tradition front and center.

A good *arja* drummer is a master of nuance. Within the confines of a metric structure laid out by gongs and other cycle-marking instruments, he improvises complex patterns to complement those of his simultaneously improvising partner. As we saw in Video 1, this often happens at a break-neck pace of over 800 drum strokes per minute. I wanted to understand how Pak Tama and other drummers like him were able to interlock while improvising. And I wanted to understand how I, as an *arja* drummer, could create new *lanang* patterns that interlocked effectively with my *wadon* partner. Yet as in many aural traditions, *kendang arja* does not have a detailed or explicit music theory. Aspiring *kendang* players most often learn by watching skilled drummers play at speed, and, even in more formalized lesson settings, there is little if any technical discussion of the music. This informal learning process has led to a largely tacit knowledge of the practice (see [3: 34–39, 10]). Information comes out unpredictably, over time, and like the statement "the *lanang* is the bus driver," often requires translation.

Knowing how to interpret the things we transcribe, then, demands a multi-tiered research approach combining transcription and analysis with extensive listening, performance experience, lessons and interviews, cultural immersion, and long conversations with musicians. In my analyses of *kendang arja* improvisation, knowledge of related Balinese genres would give me a broad understanding of basic principles of Balinese music and Balinese interlocking. Performing experience in *kendang arja* would help me understand the physical and aesthetic challenges of the practice, embodying musical principles alongside technique. And lessons with drummers across Bali would give me a collection of consciously known *arja* patterns: patterns upon which, I hypothesized, drummers might base their improvisations. Yet it would be the verbalized assertions during a lesson or the off-hand comments offered over a cup of sweet Balinese coffee that would truly shape my understanding of the practice.

3 An Introduction to *Kendang Arja*

Drums, *kendang*, are central to many Balinese *gamelan* genres, their interlocking rhythms directing ensembles through shifts in structure, tempo, and dynamics. In the island's more internationally renowned *gamelan gong kebyar* ensemble (see [19]), the elaborate patterns of the drumming pair are often lost to the novice listener behind a mass of ringing bronze. But the pared-down *geguntangan* ensemble used in the dance-drama *arja*, which Western observers once called "Balinese opera," puts the interlocking *kendang* center stage. This pair of musicians leads the ensemble of small gongs and cycle-marking instruments in Fig. 2 along with a cast of singer-dancers accompanied heterophonically by bamboo flutes.[1]

[1] On this and other general aspects of *arja* performance, see Dibia [8], Widjaja [23], Tilley [20: Chap. 1].

Fig. 2 *Gamelan geguntangan*: *kendang* and cycle-marking instruments for *arja* (Photo by Chelsea Edwardson, 2011. Used with permission)

	High (Rim) Stroke	Low (Bass) Stroke
***Wadon* (lower)**	*kom* (o)	*Dag* (D)
***Lanang* (higher)**	*peng* (e)	*Tut* (T)

Fig. 3 Main drum strokes in *kendang arja*

3.1 The Drum Strokes

Arja drum patterns are made through a blend of bass and ringing strokes of various pitches interspersed with softer unpitched finger taps. For the purposes of this chapter, four main pitched strokes are relevant.

Per Fig. 3, each drum in the pair has a high ringing stroke, called *kom* on the lower *wadon* and *peng* on the higher *lanang*. Following their Balinese mnemonics, I abbreviate these strokes "o" and "e" respectively.[2] Both are played with the left fingertips on the rim of the drum's smaller head, as shown in the photographs in Fig. 4.

Each drum also has a bass stroke played with the right thumb in the center of the larger drumhead, per Fig. 5. This stroke is called *Dag* (D) on the *wadon* and *Tut* (T) on the *lanang*. In *kendang* performance practice, *wadon* drummers often play two thumb strokes consecutively. Balinese musicians consider the first stroke in these "double *Dag*" to be subordinate to the second, aesthetically similar to a grace note.

To kinesthetically maintain running 16th-notes at almost all times, when an *arja* drummer is not playing one of her main strokes, she will lightly tap softer subordinate strokes on one of the two drumheads. The primary reason for this idiom is a practical one: when interlocking at such high speeds, physically marking the pulsation allows

[2]Both Asnawa [1] and Hood [11, 12] use 'k' and 'p' to represent these rim strokes. I use 'o' and 'e' to avoid confusing them with the common Balinese slap strokes *Kap* (K) and *Pak* (P).

Fig. 4 Playing technique for *kom*/*peng*. I Wayan Tama (left) and I Cokorda Alit Hendrawan (right) (Photos by Chelsea Edwardson, 2011. Used with permission)

Fig. 5 Playing technique for *Dag*/*Tut*. I Dewa Nyoman Sura (left) and I Cokorda Alit Hendrawan (right) (Photos by Chelsea Edwardson, 2011. Used with permission)

drummers to stay in time, and rests are thus virtually nonexistent in the practice. Some Balinese musicians term these softer subsidiary strokes *anak pukulan*: "child strokes." They may also be thought of as "ghost strokes," a term used to discuss softer strokes in Western drumset playing as well as in Mande and Northern Sotho drumming (see [4, 15, 22]). After Hood [11, 12], I refer to an *arja* drummer's subsidiary strokes as "counting strokes."

3.1.1 The Transcriptions

Most figures in this chapter are transcribed using both Western staff notation and letters to represent Balinese mnemonics. Music for *arja* is cyclic and, true to a Balinese conception of meter as end-weighted (with strong beats marking ends of cycles rather than their beginnings), the strong beat at the beginning of each transcribed pattern is shown in parentheses. This indicates that it actually belongs to the previous cycle. Main strokes are shown as round noteheads, with *lanang* strokes higher than their parallel strokes on *wadon*. *Kom* (o) and *peng* (e) are shown high on the staff, *Dag*

(D) and *Tut* (T), low. Finger-tap counting strokes are notated with [x]-noteheads in the center of the staff and "_" in mnemonic notation, while the *wadon*'s subsidiary strokes played with *Dag* technique, also occasionally used by *lanang*, are notated with low [x]-noteheads and either "d" or "t." *Kendang* patterns, which always subdivide four notes to the beat, are transcribed in 16th-notes. For visual clarity, numbers above notations indicate beats in the cycle, dots mark the half beat, and dashes delineate the other quarter-beat or 16th-note subdivisions. My hope is that this hybrid approach to transcription will ease understanding for those versed in staff notation while also encouraging engagement by Balinese musicians who may not be.

3.2 The Basics of Kendang *Interlocking*

When playing or composing interlocking patterns for Balinese *kendang* of any genre, more often than not like strokes are paired together: *kom* (o) with *peng* (e), *Dag* (D) with *Tut* (T). Figure 6 shows a simple example of this kind of strict interlocking in a 4-beat pattern, with *lanang* on the top staff, *wadon* in the middle, and the composite of their main strokes notated on the bottom. Note that though this particular pattern involves a strict alternation between *lanang* and *wadon* main strokes, this will not always be the case.

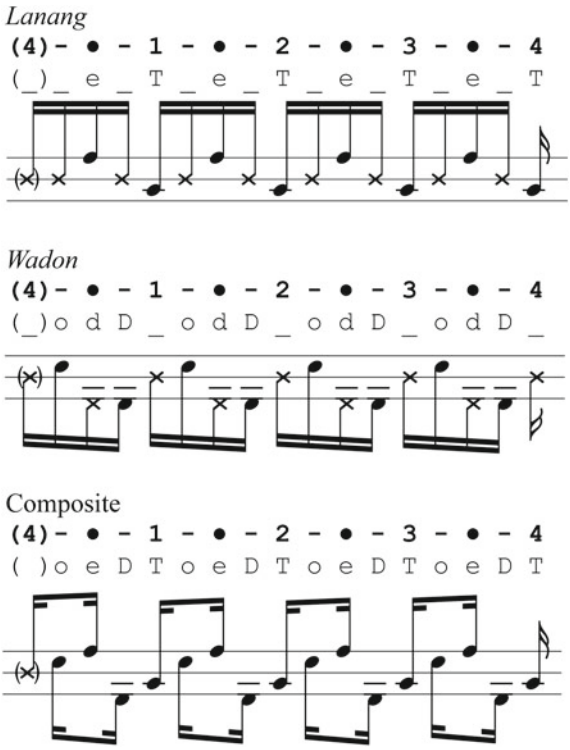


Fig. 6 Simple interlocking pattern

The pattern in Fig. 6 is simple to play, but only because the passage is pre-composed, and both musicians would have been taught their rhythm as part of the interlocking whole. In *kendang arja*, by contrast, interlocking patterns are created through simultaneous improvisation. No surprise that *arja* drummers are among Bali's most respected musicians. Audio tracks 1 through 3 will give the reader a sense of the complexity of this process. They are excerpts from a recording session with the drummers featured in Video 1, also shown in the photograph in Fig. 1: I Cokorda Alit Hendrawan (Cok Alit) of Peliatan on *lanang* and I Dewa Nyoman Sura (Pak Dewa) of Pengosekan on *wadon*. The two are long-time drum partners, and I recorded them playing together in the summer of 2011. Each of the three tracks is accompanied by *arja*'s cycle-marking instruments. A struck wooden or bamboo beat-keeper called *guntang*, quite hard to hear in this recording, marks a beat every four drumstrokes, while a small high-pitched gong called *klenang* emphasizes every second beat. A low-pitched instrument called *gong pulu* plays every 8 beats to mark the end of each cycle, and a mid-range gong called *tawa-tawa* plays at the cycle's midpoint. All but the *klenang* can also be seen in Video 1. Together these four instruments outline an 8-beat cyclic structure called *tabuh telu*, which is the longest and slowest of four structures used in *arja* performance, contrasting the slightly faster 4-beat structure from the video. Track 1 features just the higher *kendang lanang* isolated for ease of listening [Audio 1]. Track 2 highlights the lower *kendang wadon* [Audio 2].

In track 3 we hear both drummers playing the excerpt together. The reader should listen for the interlocking of high ringing strokes, light unpitched taps and, perhaps easiest to decipher on first listening, the bass-y *Dag* (D) and *Tut* (T) strokes. These strokes tease around one another without colliding, yet their interactions are not pre-composed. Each musician plays, in no pre-set order, a large variety of different patterns, both those that are dense with bass strokes and those that are sparse.

The transcription in Fig. 7 shows mnemonics for just the *Dag* (D) and *Tut* (T) strokes from the first eight cycles of the excerpt in Track 3. Here each line represents one cycle of music, the last subdivision of each shown again in parentheses at the beginning of the following line. In this excerpt, sometimes a *Dag*-heavy pattern will line up with a *Tut*-heavy one as though they had been planned in advance, but very often this is not the case. And while *Dag* (D) and *Tut* (T) strokes do occasionally coincide (circled), the musicians generally deftly interleave these strokes. This is improvised interlocking [Audio 3].

So how are they doing it? Simply asking my teachers about their techniques yielded little. When pressed, each gave me his own version of the same assertion: "we just play whatever we want. You can too." My own dismal attempts at improvisation told me this was not a statement to be taken literally. Yet, through months of lessons and casual conversations, certain other information began to trickle out: guiding principles that appeared to govern improvised interlocking.

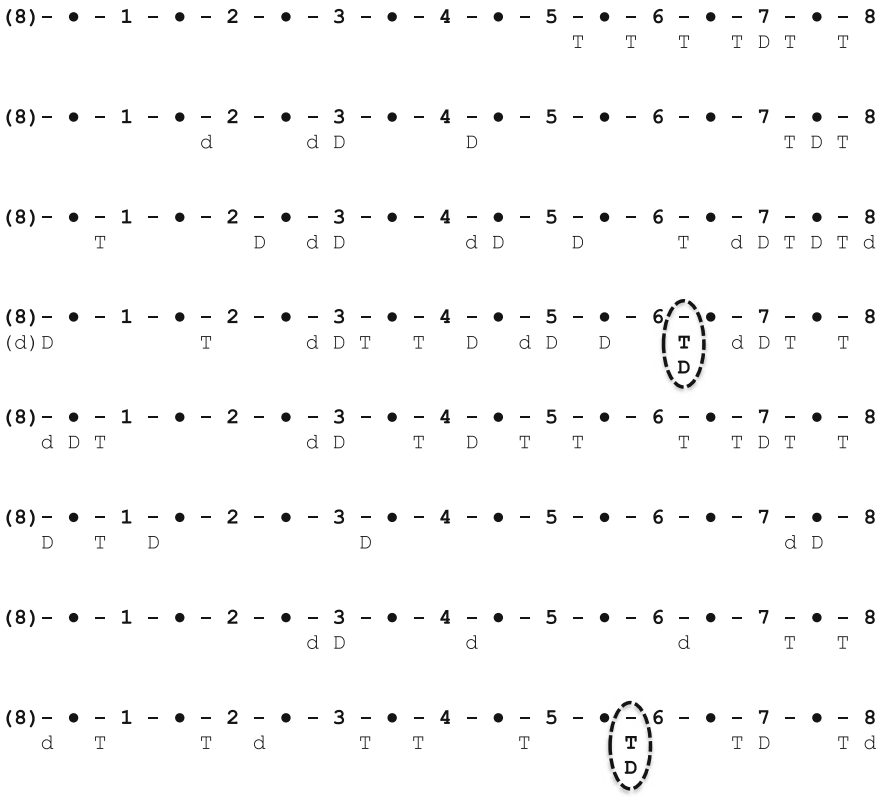


Fig. 7 *Arja* interlocking is not pre-composed

4 The On-Beat-Off-Beat Guideline

Among my many teachers and drumming friends, probably the most universally articulated guideline for *arja* improvising is that the *lanang* should play on the beat while the *wadon* plays off the beat. Different teachers revealed this directive to me in quite dissimilar ways. I Wayan Tama (Pak Tama) of Singapadu used the terms *ganjil* and *genap*, "odd" and "even," to discuss drum stroke placement. Cok Alit of Peliatan, though we spoke entirely in Indonesian, actually used the English-language terms "on-beat" and "off-beat." Some drummers and local scholars talked about the *lanang* in relation to the *mat*: the stroke of the time-keeping instrument *guntang* (Fig. 8), which, like a flexible metronome articulates every fourth drum stroke, transcribed here as quarter-notes in a simple meter. To these musicians and academics, the role of the *lanang* in *arja* was to *ngematin*: to emphasize the beat, or *mat*.

Despite differences in terminology, then, there appeared to be fairly unanimous consensus: the *lanang* played on the beat and the *wadon* played off the beat. It seemed a clear-cut rule. But as noted above, in *kendang arja* both players' hands tend to strike their drums continuously, playing either a main stroke or a counting stroke every 16th-

Fig. 8 The *guntang* beat-keeper (Photo by Chelsea Edwardson, 2011. Used with permission)

note. So what did the guideline mean? When asked to clarify, most drummers would simply play a pattern and then triumphantly say, "See?" Clearly they saw something I didn't. Yet no matter how I phrased the question, and no matter how frequently I tried to steer a conversation to the topic, none of my teachers had a response that satisfied my craving for theoretical understanding. None but I Gusti Nyoman Darta (Komin), a young Pengosekan-based drummer so curious about his own music that he had taught himself to play left-handed in order to know explicitly what other drummers implicitly knew (but could not describe) about playing technique. In a lesson one rainy August afternoon, he unceremoniously informed me, completely unsolicited (and to my great delight), that the most basic composite pattern for *kendang arja*, the *dasar*, was conceptually the pattern in Fig. 9.

In this theoretical *dasar*, the *lanang* strikes *peng* (e) "on the beat," aligning with the *guntang* time-keeper as well as the half-beat between each *guntang* stroke, while the *wadon's kom* (o) strokes occupy the "off-beat" 16th-notes (or quarter-beats) in between. Approaching my analyses with Komin's *dasar* in mind, I found that improvised patterns did often proceed, broadly speaking, with this sort of stroke placement. The on-beat-off-beat guideline, then, seemed to apply to high rim stroke

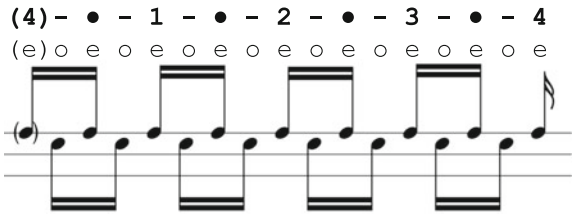


Fig. 9 Most basic composite pattern (*dasar*) for *arja*

use, *arja*'s most basic composite pattern being a strict alternation between off-beat *kom* (o) strokes and on-beat *peng* (e) strokes.

Of course no experienced drumming pair would play a pattern this repetitive for very long. And there are several different ways that drummers vary this simple pattern while still abiding by the on-beat-off-beat mandate. We know that *arja* drummers generally prefer to maintain continuous motion when playing, lightly tapping "counting strokes" when not playing one of their main strokes. In the basic *arja* pattern from Fig. 9, where each drummer appears to rest between left-hand rim strokes, he is in fact filling in the spaces with a counting stroke in the right hand. The two components of the basic pattern in Fig. 10 are shown with asterisks and [x]-noteheads denoting each drummer's right-hand counting strokes.

The simplest variation on the *dasar* replaces a main left-hand rim stroke—a *kom* (o) or *peng* (e)—with a counting stroke in the *left* hand, creating variety while still preserving the one-to-one right-left alternation that makes this pattern so basic. In Fig. 11, the first transcription is of the basic *wadon*. The second and third show two typical improvised *wadon* variants, their right-hand counting strokes now transcribed as underscores and [x]-noteheads while the left are marked with asterisks in mnemonic notation and circled in staff notation. In the first of these two variants, counting strokes replace main strokes symmetrically, every other left-hand stroke; in the second, the combination of main and counting strokes is asymmetrical.

Each of these variants creates its own unique composite with the basic *lanang*, per Fig. 12.

Because the *lanang* player, too, can replace any of his *peng* (e) strokes with a counting stroke, we get a total of 65,536 (or 2^{16}) different possible versions of this seemingly simple 4-beat phrase. Exponential growth means that, at 8 beats long, the same *dasar* has 4,294,967,296 (or 2^{32}) possible realizations. And, while individual drummers often stick to a handful of favorites, each of these possibilities can be, and most probably has been played by an *arja* drummer somewhere at some point.

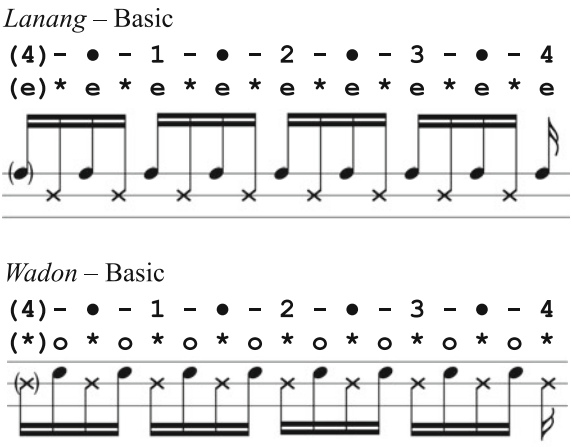


Fig. 10 Basic pattern with right-hand counting strokes (*)

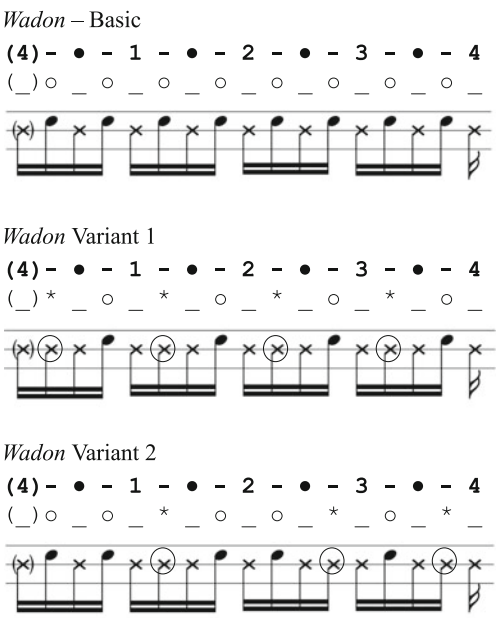


Fig. 11 Basic *wadon* (top) and two counting-stroke variants

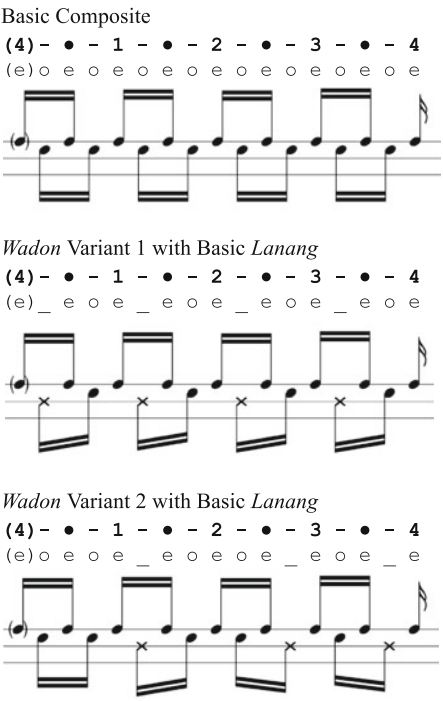


Fig. 12 *Wadon* variants from Fig. 11 paired with basic *lanang*

4.1 Double Strokes

Things become even more interesting when drummers play two rim strokes in a row, disrupting the regular right-left alternation of the *dasar*.[3] When these common double strokes occur, the "on-beat-off-beat" guideline appears to get turned on its head. Where *lanang* lines up with the slower beat levels in the metrical grid when using single strokes as we saw in Fig. 9, when double strokes are used, it is the *wadon* that lands on the beat. The two theoretical *dasar*, or basic patterns, of this more nuanced rule set are shown in Fig. 13, their on-beat strokes marked with circles and arrows.

Most *arja* drummers seamlessly blend single and double rim strokes in quick succession, alternating between these two aspects of the on-beat-off-beat guideline. In Fig. 14, the two single *peng* (e) strokes near the end of the *lanang* pattern fall on the beat, while the *wadon*'s single *kom* (o) strokes before beats 1 and 3 both land off the beat. Yet each pattern also employs double rim strokes, and these reverse on-beat-off-beat roles at the slower beat level. Together the two patterns create a composite that features the *wadon*'s double *kom* (o) strokes firmly "on the beat" (circled in black). Here and elsewhere, "l" and triangular noteheads indicate a slightly more prominent left-hand counting stroke that creates a soft slap. (For visual clarity in Figs. 14 and 15, only main rim strokes *kom* (o) and *peng* (e) are shown in the composite patterns.) Subdivisions where both drummers play counting strokes, including soft "l" slaps, are notated with rests in staff notation and underscores in mnemonic notation.

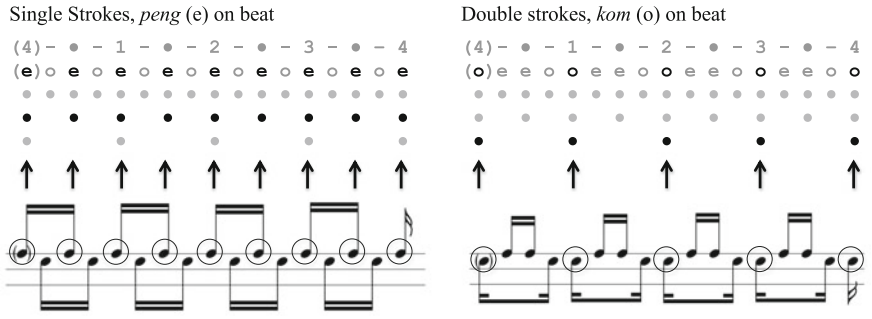


Fig. 13 The on-beat-off-beat guideline: single versus double strokes

[3]Like with the double *Dag* [d D] discussed above, the second of two rim strokes is almost always considered the stronger one, likely an artifact of end-weighted thinking.

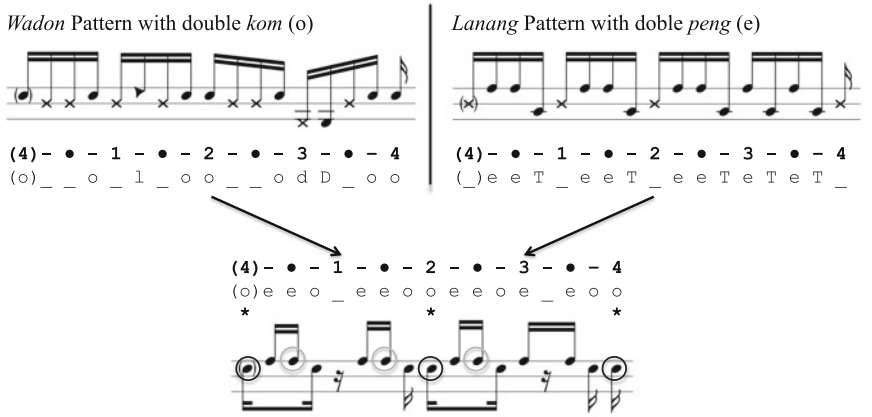


Fig. 14 The on-beat-off-beat guideline in improvisation. *Lanang* (left), *wadon* (right), composite (bottom)

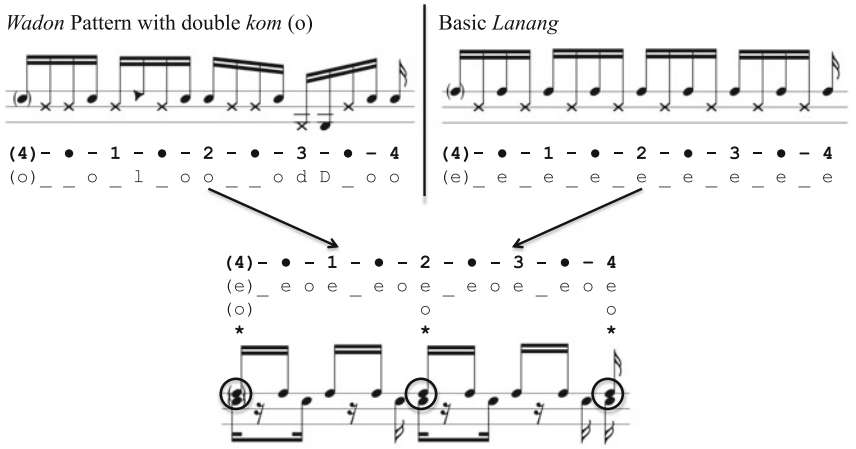


Fig. 15 Guidelines for single and double strokes contradict, causing collisions

Of course this contradictory rule set between single strokes and double strokes creates occasion for like strokes to collide instead of teasing around one another. Figure 15 shows a theoretical combination between the *wadon* pattern from Fig. 14 and the basic single-stroke *lanang*. In three places, all aligning with major beats, there is a collision of simultaneous rim strokes. Asterisks and circles in the composite pattern mark each one.

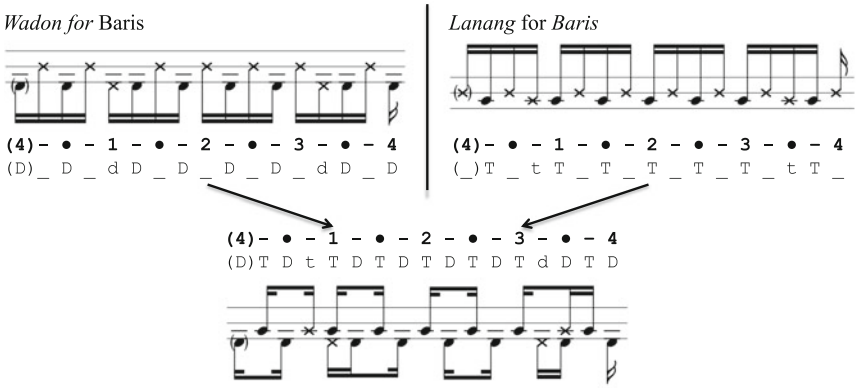


Fig. 16 Interlocking drumming for *Baris* is pre-composed

We might imagine this combination of patterns to be a faux-pas in Balinese interlocking, where like strokes are designed to dance around one another, never colliding. But when I asked Cok Alit about it, he laughed and exclaimed, "This isn't *Baris*!"

What he meant was that *arja* isn't pre-composed, like the drumming for the famous dance piece *Baris* shown in Fig. 16. It's not meant to have perfectly interlocking drumming with no collisions. As long as there weren't too many collisions in a row, drummers listening to recordings of their own *arja* playing would still be satisfied with a collision here and there. In fact most musicians from the old generation lamented that the *arja* of the young generation had become too "perfect." What was important for these drummers was that the overall impression was one of interlocking; especially at such high speeds, the details were sometimes incidental and small collisions of *kom* (o) and *peng* (e) could be enjoyed, even wished for. They reminded the listener that the music wasn't pre-composed, drawing attention through their imperfections to the near-perfection of the interaction.

4.2 The Bass Strokes

Details became significantly more important, however, and collisions far less common or seemingly acceptable, when I began looking at the bass strokes: the *wadon*'s *Dag* (D) and *lanang*'s *Tut* (T). This greater attention to detail is hardly surprising; among Balinese musicians and dancers, *Dag* (D) and *Tut* (T) are universally understood to be more structurally significant than *kom* (o) and *peng* (e). In all Balinese genres that use *arja*'s collection of drumstrokes, including the famous dances for *legong* and *gamelan gambuh*, major dance movements most frequently coincide with *Dag* (D) and *Tut* (T). Thus the smooth interaction of bass strokes is key to a satisfying performance, even in improvised *arja*. Some *arja* drummers will go so far as to make the *Dag* (D) and *Tut* (T) elements of their playing entirely pre-composed,

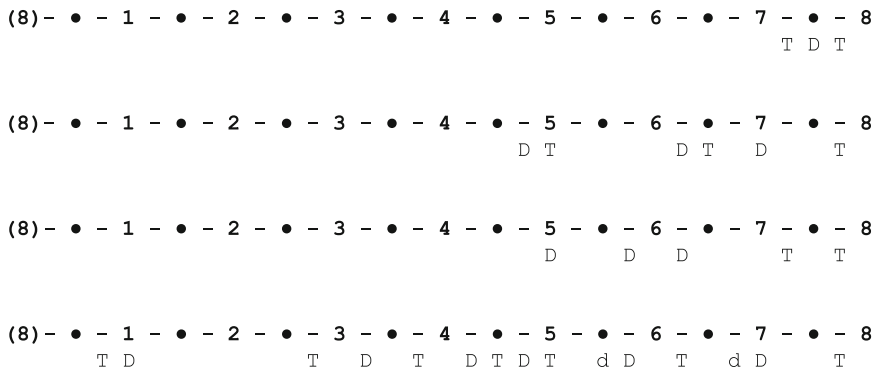


Fig. 17 Pak Tut's fixed *Dag-Tut* interactions for *arja*

so as to avoid any unwanted collisions. The drummers of Apuan village in Bangli regency, led by I Ketut Bicuh (Pak Tut), always cycle through the same four patterns, always in the same order. While they use much of the *kom* (o) and *peng* (e) variation just explored, placement of *Dag* (D) and *Tut* (T) is utterly fixed, per the mnemonic notation in Fig. 17. Audio track 4 cycles through these four patterns twice [Audio 4].

While this sort of pre-composition in *arja* is relatively uncommon, all drummers' placement of *Dag* (D) and *Tut* (T), whether pre-composed or improvised, appears to be significantly more fixed than *kom* (o) and *peng* (e). *Lanang* player Cok Alit, for instance, never places a *Tut* (T) stroke directly on the beat. And only in very rare, controlled situations does he place one on the half-beat. As we can see in two of his favored patterns in Fig. 18, Cok Alit plays *Tut* (T) strokes on the 4th subdivision of a beat about 75% of the time (circled in black) and on the 2nd subdivision about 20% (circled in grey).

This meticulous *Tut* (T) placement allows his *wadon* partner Pak Dewa to confidently play *Dag* (D) on any beat or half-beat with no worries of a collision. And analysis of the latter's patterns shows that many of his *Dag* (D) strokes do, in fact, occupy these "on-beat" positions, as we can see in the two examples in Fig. 19.

However, Pak Dewa's patterns also exhibit a freer use of *Dag* (D). In the pattern in Fig. 20, while *Dag* (D) strokes primarily occupy their "allotted" positions (shown with dashed rectangles), two *Dag* (D) marked with circles and arrows land on the second subdivision of a beat, purportedly a spot reserved for *lanang* use.

This sort of incursion into *lanang* territory is not a one-off; *wadon* players across village styles seemed to "break the rules," risking collision in their bass strokes, far more often than *lanang* players. And here we return to this chapter's scene-setting opening, and that strange expression: "the *lanang* is the bus driver."

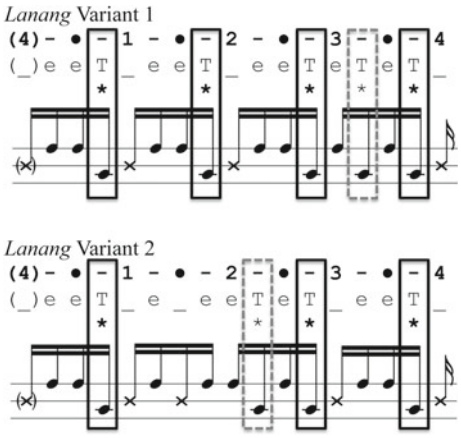


Fig. 18 Cok Alit's *Tut* (T) placement

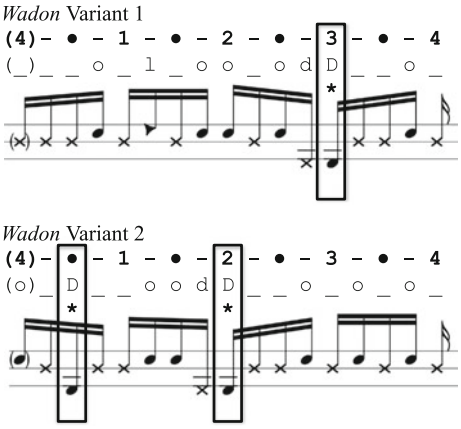


Fig. 19 Pak Dewa's on-beat *Dag* (D) placement

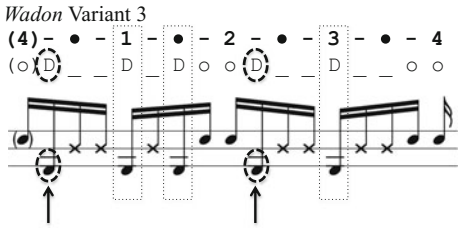


Fig. 20 Pak Dewa's freer use of *Dag* in off-beat positions

5 Instrument Roles

It came out during one of my lessons with Pak Tama of Singapadu, as most of his best information came out: with no prodding from me whatsoever. "You see," he begins, "the *lanang* is the leader. He has to communicate cues from the singer-dancer to the ensemble." This was something I already knew; it was one of the reasons I was so

terrified to play *lanang* in the temple. The *lanang* had to be able to stop on a dime, interrupt any pattern she was currently improvising to give cues to the ensemble at the whim of the singer. Pak Tama compares the relationship between *lanang* and *wadon* to that between *sopir* and *kernet*: the bus driver and the helper who puts the passengers' bags on the top of the bus. The *lanang* player is the *sopir*; he is in control of the bus. He must constantly be watching the road, following it carefully. He cannot stray off the road, or play around too much lest the bus tip over. The *wadon* player is the *kernet*: he has more freedom to move around the bus, placing bags wherever he finds space and socializing with the passengers. Pak Tama dances his hands back and forth to demonstrate the concept like the two dancers inside a *barong* costume: the front hand very carefully traces a path, while the back hand moves around more freely, ducking left and right, sometimes coming very close to the front hand then backing away playfully.[4] To Tama, the *wadon* is the drum that makes the sound exciting: "We play like this so that it sounds like the voice of the drums is alive," he says. "The *wadon* is what makes it come alive. Yeah, here [the *lanang*] always continues in the same way. The *lanang* has some variations too, but far fewer."[5]

Indeed, many studies of group improvised musics show that in order for one musician to have ample improvisatory freedom, another must provide a framework of stability. We see this in the balancing of soloist and rhythm section in jazz [2, 16] or lead and supporting drummers in various Ghanaian traditions [5, 13], as well as in the more stable *kushaura* and freer *kutsinhira* improvisations among many *mbira* performers (personal communications). Other *arja* drummers make similar observations about their own playing, and I have certainly found it to be true in my analyses as well. In fact, one of the first things that I Wayan Sudirana (Sudi) of Ubud taught me was that just a few short, basic, *lanang* patterns could be paired idiomatically with virtually any *wadon* pattern, no matter how long or complex. Like most of my teachers, the *lanang* patterns he taught me were all four or eight beats long. What's more, all appeared to be small variants the one upon the other. In the two taught patterns in Fig. 21, the second is simply a pared-down version of the first, cutting the *Tut* (T) strokes and the second of each double *peng* (e).

The *wadon* patterns that Sudi taught, by contrast, were often eight, twelve, even up to twenty beats long. These patterns, in both length and content, were much more varied and complex than the accompanying *lanang* patterns. The same was true of each of the other teachers from whom I learned patterns for both drums. Although *lanang* players like Cok Alit do play a wide variety of patterns, they are generally shorter, less varied, and less complex than their partners' accompanying

[4]The Barong is a mythical lion-shaped demon featured in, among other things, the *Calonarang* story, where he fights the widow-witch Rangda in an epic battle of good against evil. Like the Chinese dragon, the Barong is played by two dancers in a single costume performing a sometimes comical and oft-beautiful dance. The dancer performing the front legs also controls the mask and its snapping jaws, while the one in the back controls the rear legs and rump.

[5]"Kita main begitu supaya kedengaran suara *kendang* itu hidup. *Wadon* yang hidupnya. Ya di sini [di *lanang*] tetap jalan. […] *Lanang*nya ada juga variasi sedikit, tapi lebih sedikit disini" (Conversation, June 2011).

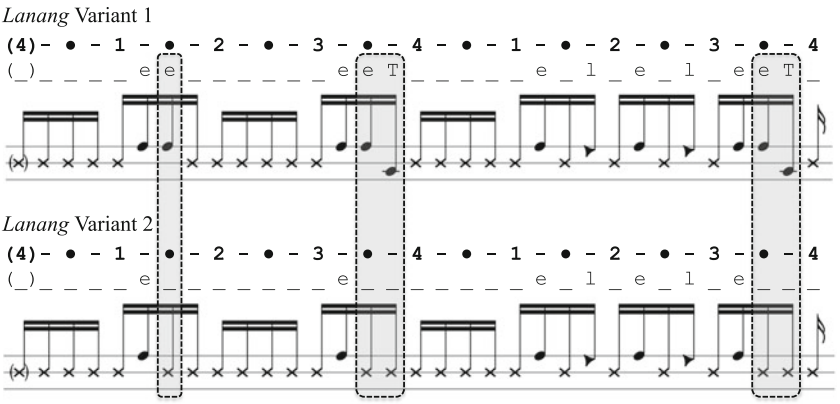


Fig. 21 Sudi's *lanang* patterns as close variants of one another

wadon patterns. In other words, improvisations on the *lanang* should be relatively simple, allowing the *wadon* more freedom to take risks.

With that in mind, we return to the relatively free use of *Dag* (D) from Fig. 20. In this pattern, Pak Dewa plays two of his *Dag* (D) strokes, which we would expect either on the beat or the half-beat, instead on the beat's 2nd subdivision. In doing so, he risks collision with his *lanang* partner. Yet, as we saw in Fig. 18, the majority of Cok Alit's *Tut* (T) strokes land not here but in a beat's 4th subdivision. Pak Dewa, then, is still fairly safe in this *Dag* (D) placement. Further, most of Cok Alit's 2nd-subdivision *Tut* (T) strokes occur late in a cyclic structure, as in the patterns in Fig. 22.

Pak Dewa, who usually restricts his off-beat *Dag* (D) use to earlier beats in a cycle, is thus still quite unlikely to crash into his partner. The relative rigidity of Cok Alit's *Tut* (T) idioms, and Pak Dewa's familiarity with them, allows the latter increased *Dag* (D) flexibility.

5.1 Aesthetic Guidelines: Ngegongin

Alongside stroke-placement strategies are also aesthetic guidelines that help shape pattern composition. As previously mentioned, *arja*, like much Balinese music, is underscored by cycles of repeating strokes on various different gongs. The end of each cycle is marked by the most important of these gongs, and to emphasize this, Balinese music is often composed with a feeling of "leading to the gong," or *ngegongin*. Tenzer's study of *gamelan gong kebyar* examines melodies for their kinetic qualities, comparing *ngubeng*, or static, with *majalan*, or dynamic moments within them. He notes that gong strokes and other important cyclic markers "exert 'pull' or 'gravity' on melodic motion, causing regions that lead up to them to be more *majalan* [more dynamic]" [19: 179]. Many *arja* patterns likewise appear to be built to show increased activity and, particularly important, increased use of *Dag* (D) and *Tut* (T) strokes as a cycle progresses, leading to the gong.

Figure 23 shows an 8-beat *lanang* pattern from Pak Tama of Singapadu. Each quadrant of the phrase exhibits an increased use of main strokes: 3, 4, 6, and 6 respectively. What's more, particularly through the second half of the pattern, there is an increased frequency of *Tut* (T) stroke use. This densification of main strokes approaching gong exists in shorter *arja* phrases as well. Two typical 4-beat *lanang* variants from Cok Alit are shown in Fig. 24. In each, a combination of main-stroke use and *Tut* (T) density increase toward the gong stroke.

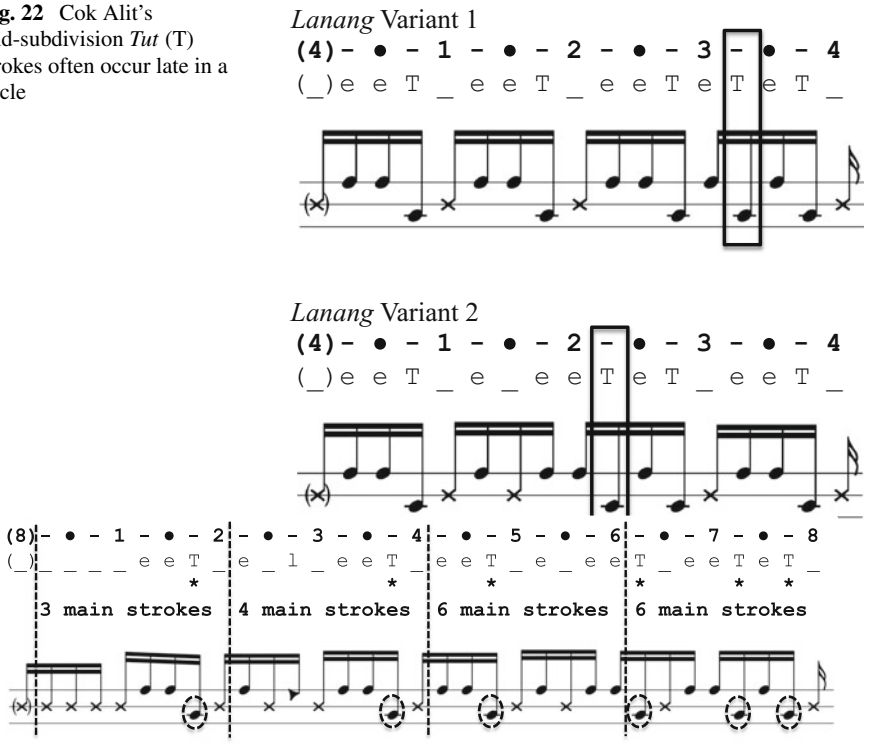


Fig. 22 Cok Alit's 2nd-subdivision *Tut* (T) strokes often occur late in a cycle

Fig. 23 Increased activity leading to gong, 8-beat pattern

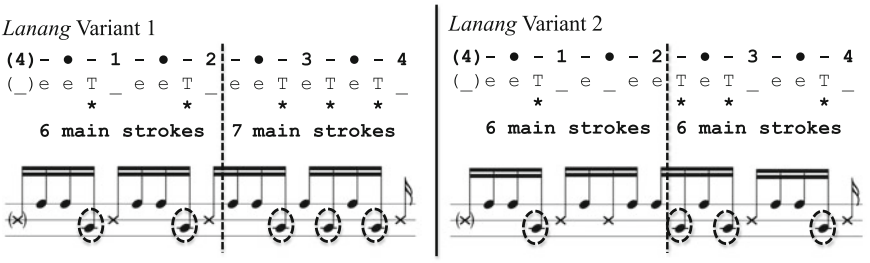


Fig. 24 Increased activity leading to gong, 4-beat patterns

5.1.1 Structural Signposting

Densification is one type of *ngegongin*, but there are others. In some *gamelan* genres, specific patterns are used to prepare structurally important moments. In the long *pengawak* cycles of the famous *legong* dances, for instance, the final gong stroke is generally preceded by a stock standard cadential phrase in the drums, recognizable as a signpost to both musicians and dancers. Figure 25 shows one such cadential formula from the *legong* drummers of Saba village. Here "K," "P," and [x]-noteheads denote loud slap strokes, round noteheads just below the center staff line indicate the *lanang*'s special cuing stroke *pung* (U), and counting strokes are marked as rests in staff notation.

In *arja*, where 2-, 4-, and 8-beat cycles are too short for extensive cadential gestures, the *wadon* player achieves this sort of "signposting *ngegongin*" simply by playing *Dag* (D) on the beat or half-beat directly preceding gong. For further emphasis, this stroke will frequently be followed by soft counting strokes or even rests (which, as we know, are extremely rare). In Fig. 26 we can see that Pak Tut's semi-composed patterns exhibit this placement, as do many of Pak Dewa's improvised ones. In ease case, regardless of cycle length, *Dag* (D) marks the beat or half-beat just before gong. It *ngegongin*s.

Of course music will seldom so obediently fit into our theoretical categories, and the concept of *ngegongin* becomes cloudy in *arja*'s relatively short cycles where drummers tend not to emphasize every gong stroke but every few cycles. In Pak Tut's patterns from Fig. 26, for instance, the third and fourth cycles are often thought of as one continuous 16-beat phrase. "Bersama," Pak Tama says of these patterns: "together." Alternately, the entire arrangement of four cycles may be thought of as a single sentence, leading through an increased density of *Dag* (D) and *Tut* (T) strokes over each subsequent cycle to a final gong at the end of the fourth cycle.

This cycle-grouping phenomenon is even more common (and complex) in shorter 2- and 4-beat cyclic structures, where there does not appear to be any fixed idea about which cycles should feature a final *Dag* (D) in improvised performance. These decisions are shaped neither by the singer nor by the structure-marking instruments, but appear to be a free choice for the drummer. Figure 27 is a short excerpt of improvised *wadon* drumming over a 2-beat cyclic structure, with strong final *Dag*s marked with asterisks and cycles grouped together to show *ngegongin* placement. It's clear that Pak Dewa does not feel obliged to *ngegongin* before each gong stroke;

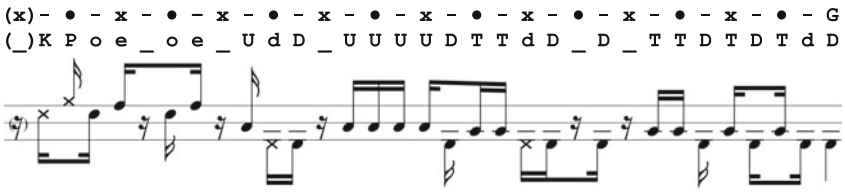


Fig. 25 Cadential formula for *legong*, Saba village style

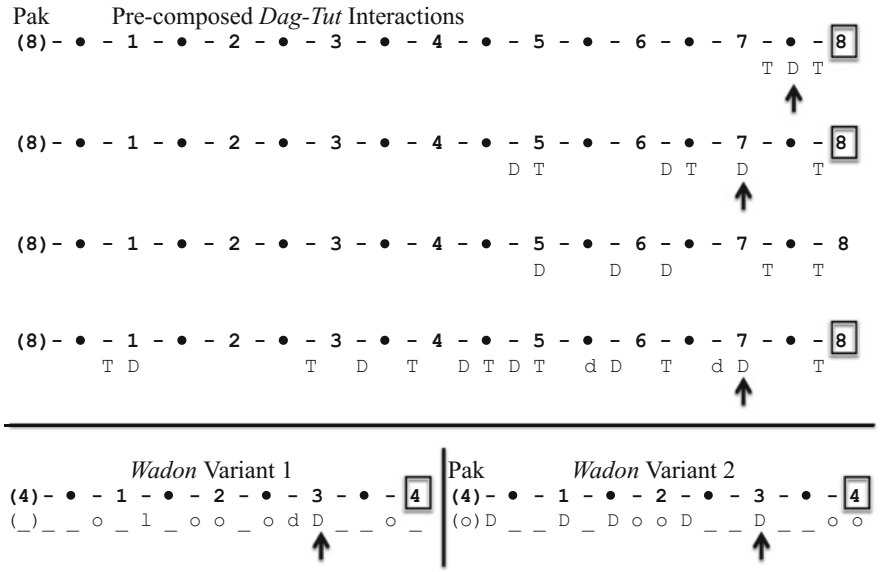


Fig. 26 Signposting *ngegongin* in 4- and 8-beat cycles

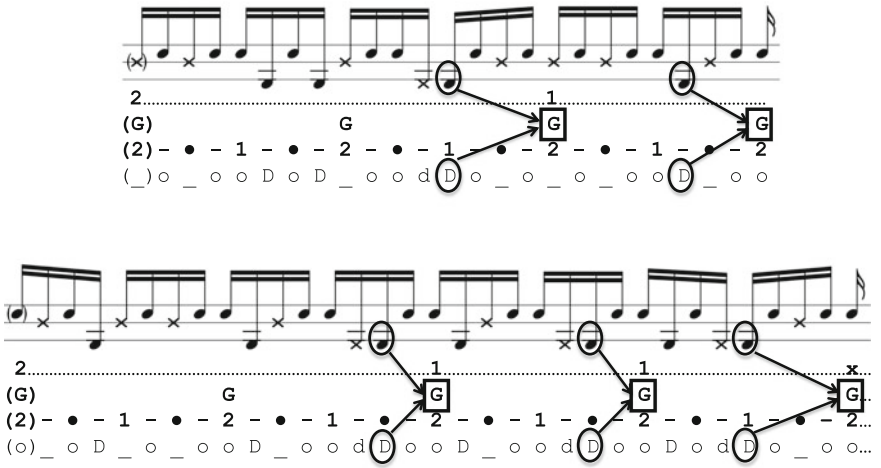


Fig. 27 Signposting *ngegongin* in 2-beat cyclic structure: improvised cycle-grouping

nor does he attempt to do so at regular intervals. We see *ngegongin* after 2 cycles, then 1, then 2, then 1, twice in a row. Sometimes he will wait three cycles; sometimes he'll *ngegongin* before each gong stroke for several cycles in a row. The decision appears to rely on little more than his whim in the moment.

6 Mixing It Up: *Campur-Campur*

The improvisation in Fig. 27 suggests a more nuanced conception of *ngegongin*, thought of not in terms of single cycles but flexible meta-cycles. And while there is no formal theory to explain such a concept, my lessons with Pak Dewa offered some corroboration. He had taught me three *wadon* patterns, two 8-beat patterns and a 12-beat one, that I was now cycling while he improvised an interlocking *lanang* part. Once satisfied that I could play these three patterns in sequence for several minutes without faltering, he upped the ante: "Boleh campur-campur," he said: you can mix them up. And then he demonstrated, playing pattern 1, then 3, then 2, then 1 again, twice. But then he started mixing them up in more complicated ways: taking the first half of pattern 1 and tacking it on to the second half of pattern 3; taking the last two beats of pattern 2 and repeating them a few times before moving on to something else.

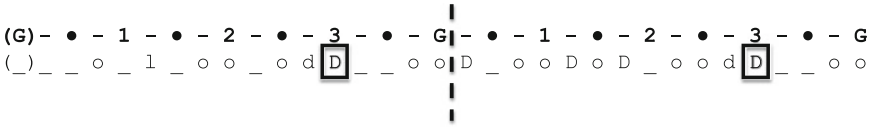
So there it was. Pak Dewa taught his patterns as 8- and 12-beat units, and he sometimes played them as such. But he conceived of them, and he used them as 2- and 4-beat musical components, to be mixed and matched at will, recombined, or played independently as the basis for improvisation. An examination of the first pattern he taught me, which he called his "basic" pattern or *dasar*, will demonstrate its usefulness in each of these different capacities. Though Pak Dewa plays this 8-beat pattern in all cyclic structures, he taught it for use in a 4-beat structure. The taught pattern is shown without analytical markings in the first transcription of Fig. 28. For visual clarity in this figure, only mnemonic notation is used. "G" represents the stroke on the *gong pulu* that marks the end of each cycle.

As we can see in the second transcription in Fig. 28, there are two signpost-style *ngegongin* in this pattern. The *Dag* (D) on beat 3 precedes and calls attention to the gong stroke at the halfway mark. This is paralleled by a *Dag* (D) four beats later, anticipating the next gong stroke. But the third transcription also reveals a more extensive densification-style *ngegongin*: the second half of the pattern has more main strokes, and more *Dag* (D) strokes specifically, than the first half, indicating a longer 8-beat trajectory. Thus, this pattern could be conceived of as a single 8-beat unit, as it was taught, but may also be felt, and used, as two 4-beat units that each *ngegongin* independently. And indeed, Pak Dewa would sometimes use this taught pattern in its entirety in improvisation; at other times he would play the first four beats without the last four, or vice versa. What's more, he would sometimes take just the first two beats and cycle them several times over, per Fig. 29.

These smaller 2- and 4-beat units, then, can be conceived of as building blocks for the longer taught patterns. And once I had discovered both the existence and the conscious use of such componential musical units, analysis became yet more exciting. The guidelines I've been discussing, whether conscious or unconscious, create musical gestures and aesthetic principles for each drum that appear to cross village boundaries, guiding my many teachers' varied patterns. In *lanang* improvising, for instance, the various permutations of the on-beat/off-beat guideline create a reality where most patterns are peppered with gestures such as "(_) e e T _" and "(e) T e T _".

Original 8-beat pattern, taught for use with a 4-beat cyclic structure:

Seen as two 4-beat patterns, using signposting *ngegongin*:



Seen as one 8-beat pattern, using densification *ngegongin*:

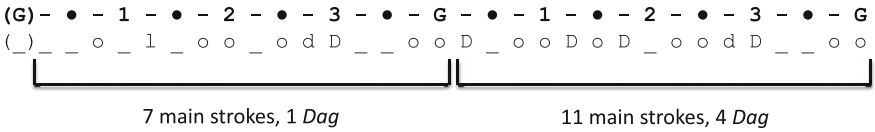


Fig. 28 Taught patterns as segmentable

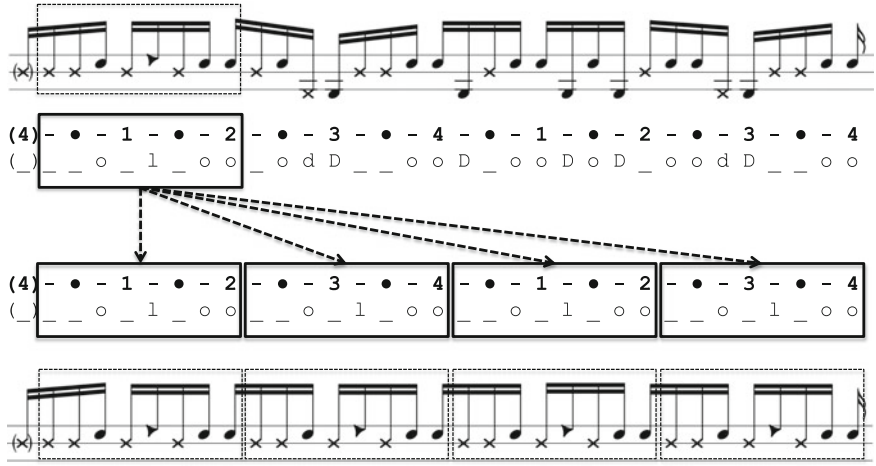


Fig. 29 Cycling a 2-beat unit

Yet these ubiquitous gestures are then fiddled with and rearranged to fashion a myriad of different patterns.

Figure 30 is a visualization of a handful of 4-beat *lanang* patterns from various different drummers, shown in this notational style to facilitate visual comparison. Each square represents one 16th-note or quarter-beat, and the patterns move horizontally. The grey squares are *peng* (e) strokes, the black, *Tut* (T) strokes. Though

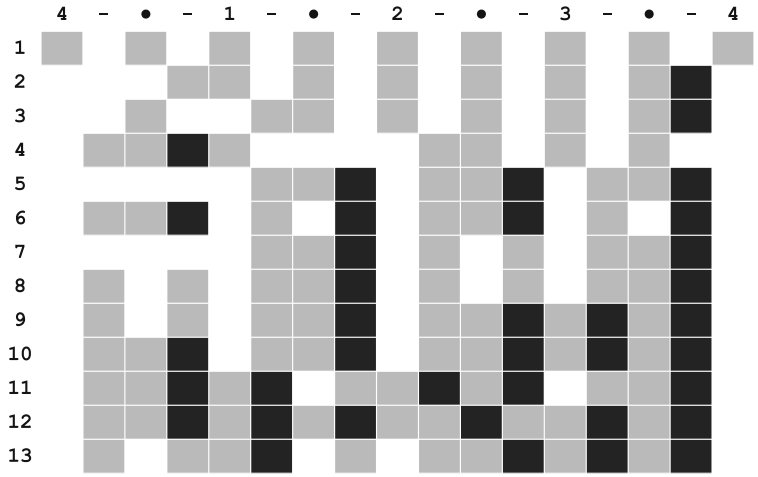


Fig. 30 Visual comparison of *lanang* patterns

all different, just a cursory glance at this chart reveals that the patterns are also, to varying degrees, rhythmically related. I hypothesized that these taught patterns, and their fundamental components, formed the basis for *arja* improvisation.

6.1 Surface-Level Variations

With every drummer I recorded, patterns played during improvisation could almost invariably be linked back to a taught pattern. But seldom in its "pure" form. Sometimes drummers vary their taught patterns at a surface level, adding extra rim strokes or, conversely, replacing rim strokes with soft counting strokes. Figure 31 is an 8-beat taught pattern from Pak Tama of Singapadu and some of the less extreme ways he varies it in performance. Here, variations from the original taught pattern are marked with dashed rectangles.

In the first variation in Fig. 31, Tama cuts the first *peng* (e); in the second, he adds an extra one. In the third variation, he adds two *peng* (e) at the beginning of the pattern. Near the end of that same pattern, he replaces one *peng* (e) with the soft slap of an "l" counting stroke.

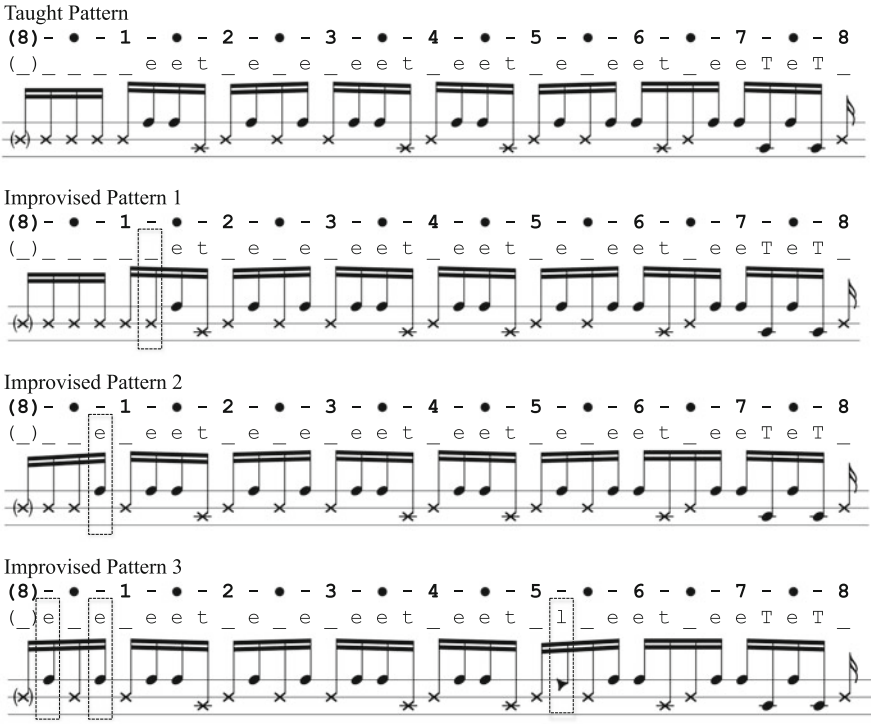


Fig. 31 Surface-level variation

6.2 More Radical Reworkings of Taught Patterns

While most *arja* drummers make extensive use of this kind of surface-level variation, they can also diverge more sharply from the taught patterns. In the analyses to follow, the reader should not be overly concerned with the details of individual notes in a long transcription. Here we seek to understand the general strategies that improvising drummers use, on the fly, to recombine familiar elements in idiomatic ways, creating new patterns that still interlock effectively with their partners' patterns. Alterations to taught patterns are marked in the figures with arrows, circles, boxes, and the juxtaposition of grey with black text.

Slightly more extreme than the surface-level variations examined above are improvisations that maintain the rhythmic identity of a taught pattern but shift the rhythm in relation to the cycle. In the two examples in Fig. 32, the actual drum strokes of the taught patterns remain intact, but are displaced by one beat, either earlier, as in Example 1, or later, as in Example 2. In these and other examples, the taught pattern is notated above the improvised one:

Related to this pattern displacement are taught patterns in which common components, like the *lanang* gestures "e e T" and "e T e T," are reversed or rearranged in

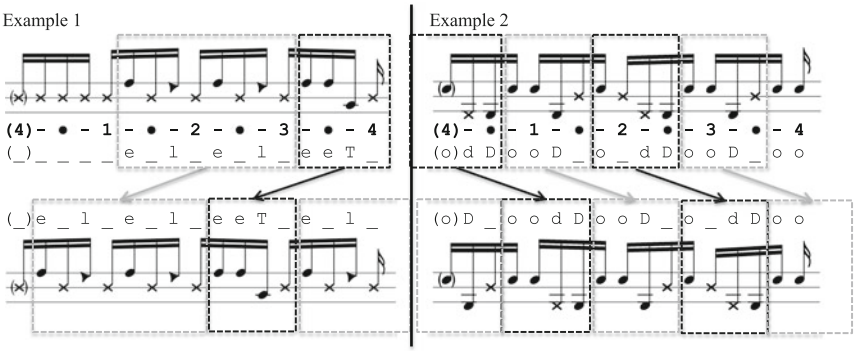


Fig. 32 Pattern displacement

improvised performance. In Fig. 33, a taught pattern and three improvised variants from Pak Tama show a simple, small-scale example of this kind of pattern element rearrangement. For visual simplicity in the next several figures, only mnemonic notation is used. Again, the reader need not get bogged down in the minutiae of each drumstroke; rather, most pertinent to the discussion are the improvisational strategies at play.

In Fig. 33, thin arrows indicate surface-level rim stroke variances, and these explain variation 1 and most of variation 2. But at the end of the second variation, something bigger happens. Pak Tama takes the "e T" and "e e T" modules from variation 1 and reverses them, giving the pattern end a different relationship to the

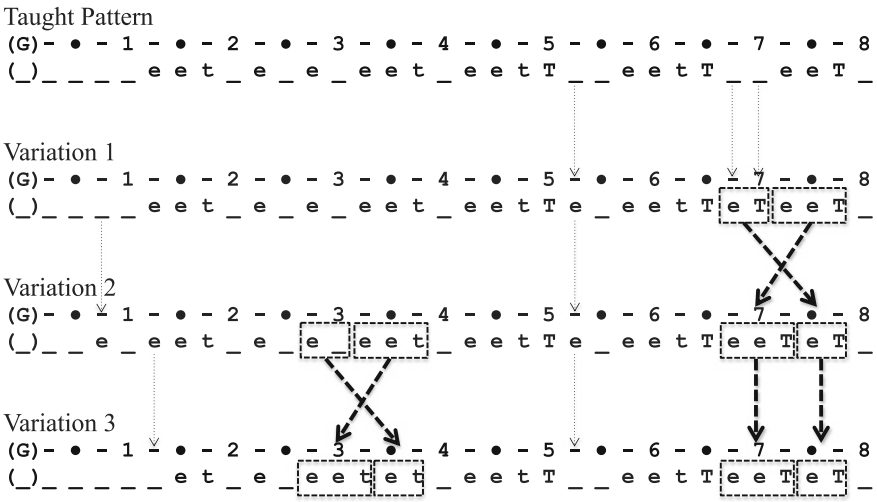


Fig. 33 Increasing rearrangement of pattern elements

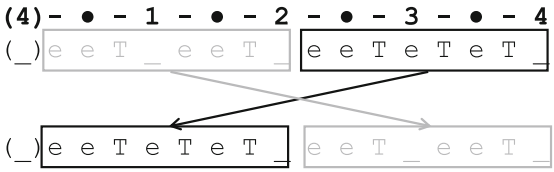


Fig. 34 Larger-scale pattern element rearrangement

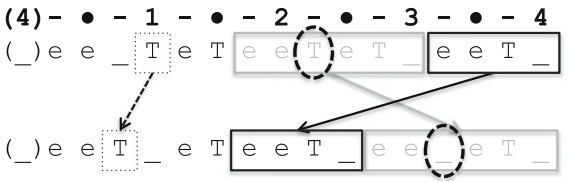


Fig. 35 More complex pattern element rearrangement

beat. Because *Tut* (T) is such a structurally important stroke, this seemingly small change actually gives the pattern a very different feeling. In variation 3, Tama also makes the same change in the middle of the cycle with his subordinate "t" strokes.

The examples in Figs. 34 and 35 show this sort of pattern element rearrangement on a larger scale. In Fig. 34, a simple reversing of the first two beats and the last two beats of the taught pattern creates a brand new pattern of the kind Pak Dewa demonstrated in my lessons.

The improvised pattern in Fig. 35 rearranges elements in a less symmetrical fashion. Here, the four-note black gesture at the taught pattern's end trades places with the 6 notes of the grey gesture so that each component now relates differently to the meter. Add to this a small shifting of the *Tut* (T) stroke at beat 1 (marked with an arrow) and the deletion of a *Tut* (T) stroke in the grey gesture (circled), and the drummer has created a very new *lanang* pattern.

Yet, different though the two patterns in Fig. 35 may be, because the second one is built from entirely familiar *lanang* elements present in the first, it is still guaranteed interlock, with an acceptable level of collision, with most any *wadon* pattern. Except where guidelines are already stretched in the original taught pattern, as in the off-beat *peng* (e) following beat 1, the new improvised pattern features single *peng* (e) strokes on the beat or half-beat and double *peng* (e) strokes mostly in their off-beat position. And though it was Pak Tama, not Cok Alit, who played this pattern, we still see *Tut* (T) strokes largely in 2nd and 4th subdivision positions.

Another kind of pattern element rearranging sees drummers take components from various different taught patterns and splice them together. Again this can be done symmetrically or asymmetrically. In the middle transcription in Fig. 36 we see an improvised *lanang* pattern in which the last two beats of one taught pattern are combined with the last two beats of another. Together they create a pattern denser in main strokes than either of the original patterns.

In the improvised pattern on the bottom of Fig. 37, the drummer begins as though he will play the taught pattern on the top left, varied at a surface level through the deletion of one *peng* (e) stroke (marked with a thin arrow). Yet his playing then

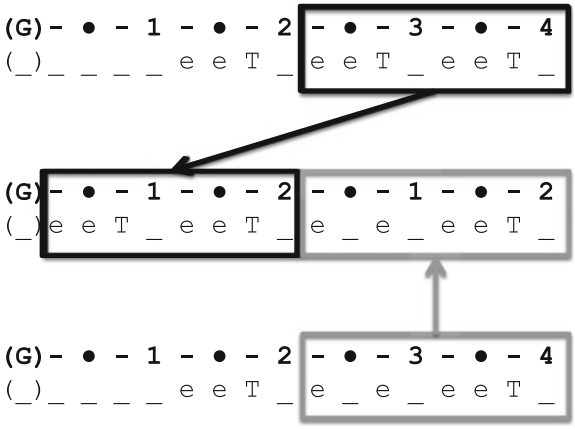


Fig. 36 Combining elements from multiple taught patterns, symmetrical

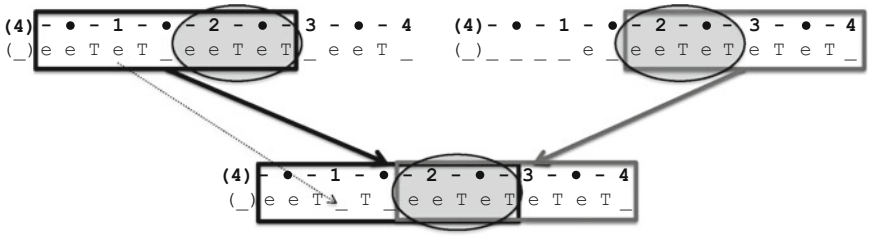


Fig. 37 Combining elements from multiple taught patterns, asymmetrical

dovetails smoothly into the ending of the taught pattern on the right through the elision of their shared "e e T e T" modules (circled in grey).

In extreme examples, such pattern splicing and reworking can lead to the creation of new cross-rhythmic patterns, an aesthetic we see often in advanced *kendang arja* performance. The transcription in the middle of Fig. 38 shows an improvised pattern built from elements of three different 4-beat taught patterns. The combination of the two taught patterns on the bottom of the figure, with some small surface-level variation, creates a 6-note motive that gets repeated 4 times in cross-rhythm to the metrical structure (circled).

6.2.1 The Risks and Subtleties of More Radical Pattern Reworking

These sorts of cross-rhythmic expansions of pattern elements can sometimes generate what seem like unidiomatic patterns. It is commonly understood, for instance, that a *wadon* player should never align his *Dag* (D) stroke with gong. The one exception to this rule is the very last gong in a piece, which will almost invariably be articulated with a *Dag* (D) stroke. In fact, one of the most common musical jokes among Balinese

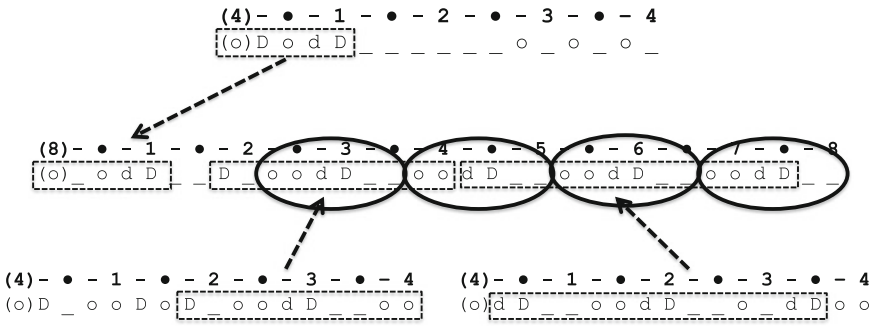


Fig. 38 Improvised pattern reworking creating cross-rhythm

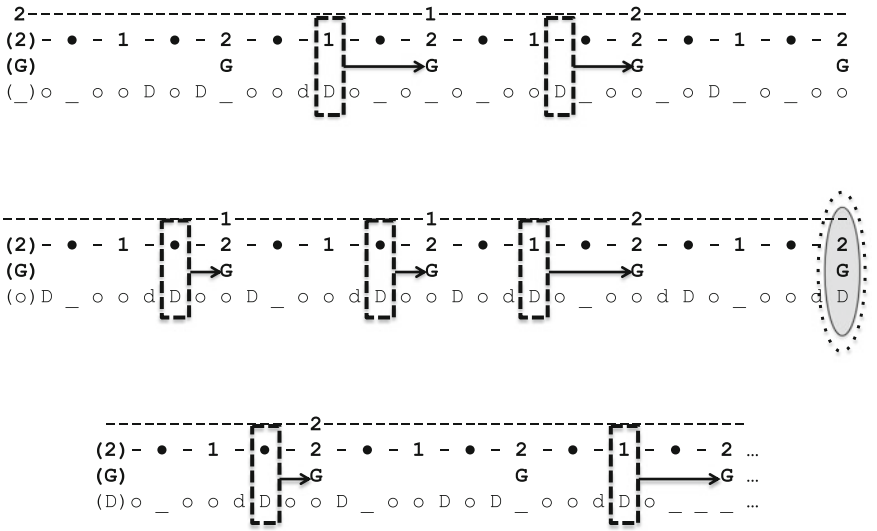


Fig. 39 *Dag* (D) on gong in improvised drumming

drummers, when playfully teasing *bulé* (non-Balinese) drummers, is to place a strong *Dag* (D) directly on gong while improvising. This will inevitably be followed by peals of good-natured laughter and wide grins in the direction of any *bulé* musician present. Yet, analysis of the improvised patterns of *wadon* drummer Pak Dewa surprisingly shows the occasional on-gong *Dag* (D) stroke. And while we could simply analyze these as mistakes (and mistakes certainly do occur), the concept of meta-cycles, explored above, allows for a more nuanced analysis. Figure 39 shows approximately 8 seconds of Pak Dewa's *wadon* improvising in a 2-beat cyclic structure. It is divided into meta-cycles according to the presence or absence of singposting *ngegongin*; prominent final *Dag* (D) strokes on the beat or half beat before gong (or, in one case, the quarter beat in between) are marked with dashed rectangles. At the end of the second line, however, the circled *Dag* (D) stroke coincides with gong.

This drum stroke is not a one-time mistake; Pak Dewa played a quite similar passage later in the same recording session. His on-gong *Dag* (D) placement, then, appears to be deliberate. Yet, while it seems to break the most unbreakable rule of all, turning this expert drummer into an awkward *bulé*, analysis of the passage in the context of its meta-cycles leads to a reinterpretation of the stroke's function. Beginning two cycles before the *Dag*-on-gong, we first see a motive from one of Pak Dewa's basic taught patterns, shown in the top transcription of Fig. 40. Aside from one surface-level variation—the deletion of a *Dag* (D) stroke—the two beats in the grey rectangles are identical. Yet instead of continuing with the second half of this taught pattern, as he does in many other improvisations, Pak Dewa shifts to the 6-note gesture "o o d D o _" and repeats it three times in cross rhythm to the cycle-marking instruments (marked with black rectangles).

Seen in this larger context, the *Dag*-on-gong (circled in Figs. 39 and 40) is simply the middle of a cross-rhythmic pattern that resolves with an appropriately-placed signposting *ngegongin* in the following cycle: a strong *Dag* (D) on the half-beat before Fig. 40's final gong. Here, flexibility of analysis, and of analytical categories, allows us to find logic in a seemingly rule-breaking passage that we might otherwise misinterpret as a simple mistake.

7 Personal Variance and Regional Style

The general principles of the oral music theory discussed above create a safety zone: a set of guidelines within which any improvised *lanang* variant will interlock effectively with virtually any improvised *wadon* variant. Yet, when I began comparing the improvisations of my various teachers, I noticed something curious. All of these drummers have a shared musical lineage, each influenced several decades ago by master musicians from the village of Singapadu, most particularly I Madé Kredek and his drum partner I Cokorda Oka Tublen. As the phylogenetic tree in Fig. 41 shows, each of my teachers learned one or both of these drummers' patterns, either directly or, in the case of Pak Tut from Apuan, as a second-generation student.

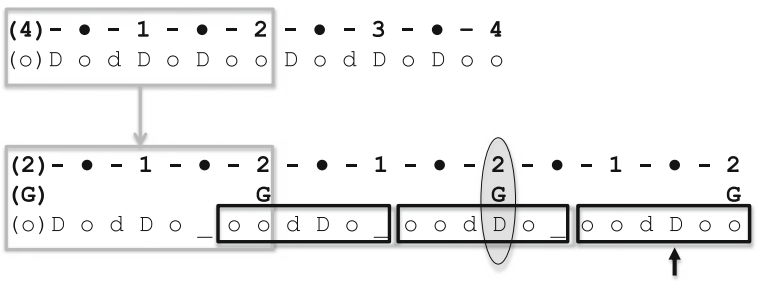


Fig. 40 *Dag*-on-gong as part of a cross-rhythm within a meta-cycle

Yet, despite their shared lineage, each of my teachers appears to have quite different levels of tolerance for variation on their taught patterns. As we saw in Fig. 17, Pak Tut never varies *Dag* (D) and *Tut* (T) stroke placement, offering only surface-level variation in his rim stroke use.

Pak Tama does vary his teachers' patterns, but usually only in quite controlled and limited ways. And he generally improvises with a narrow range of motives at any given time, these largely based on his own small collection of taught patterns. Figure 42 is an analysis of a short passage of Tama's improvised *lanang* playing. Here we are less interested in the specific notes and rhythms being used than in the larger structural logic of Tama's improvisations. Each box in the figure represents two beats, with two boxes making up a full 4-beat taught pattern and ellipses indicating a continuation of the pattern in the previous box. In this passage, Tama works with only three 4-beat patterns at once (A, B, and C), playing them in different orders, segmenting them, and varying them mostly at a surface level.

Even when using a slightly broader spectrum of patterns, Pak Tama still seems to prefer focusing primarily on just two or three at once, really taking the time to explore the nuances of each. Though the passage in Fig. 43 employs elements of five different taught patterns, Tama is mostly playing with just two, labeled B and C. By varying, segmenting, and rearranging them, and by slowly introducing other variants into the mix (X and Y), Tama keeps the overall structure unpredictable and fresh. The entire passage is then flanked by matching bookends (pattern A) that rhythmically contrast the rest of the patterns. It is an elegant but relatively conservative approach to improvisation.

All of the most extreme forms of improvisation discussed above, including the various cross-rhythmic reworkings of pattern elements, were taken from drum partners Pak Dewa and Cok Alit. Figure 44 analyses a segment of Cok Alit's *lanang* improvising, similar in length to Tama's passage from Fig. 43.

Cok Alit appears to base his improvised playing not just on his own taught patterns, but on patterns from teachers in other villages as well, implying exposure to a

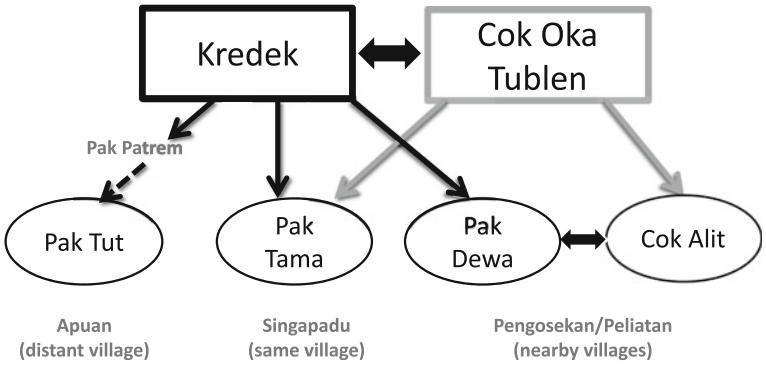


Fig. 41 Family tree of *arja* drummers

Fig. 42 Pak Tama's relatively conservative improvising

Fig. 43 Pak Tama's relatively conservative improvising stays innovative

Fig. 44 Cok Alit's more varied improvisations

much larger corpus over his lifetime than the small collection he knows consciously and teaches to his students. In Fig. 44, patterns for each different village style are represented by a different colour: five in total. What's more, while Pak Tama focuses on just a few variants at a time as we saw in Figs. 42 and 43, Cok Alit uses a large array of variants in quick succession. In the small passage in Fig. 44, he plays with 13 different variants (A–M), only one of which occurs more than once. I have written extensively elsewhere on the possible reasons for these differences in my teachers' improvisatory proclivities. Much of it I attribute to a blend of geographical location and exposure to other playing styles, as well as relative drumming proficiency level and personal ideology of preservation and change (see [21]).

7.1 Partnership

Musical analysis informed by oral music theory can go a long way to explaining the conscious and unconscious parameters guiding *kendang arja* improvisation. Yet every Balinese drummer will say that the key to success in *kendang arja*, over and above these tangible organizing principles, is *pasangan*: partnership. Long-time *kendang arja* partners develop a compatible vocabulary of patterns based on their web of musical influences and their personal opinions on drumming style and innovation. Pak Tama of Singapadu recalls being invited to play in a performance with Pak Gobleg, an accomplished *arja* drummer from the village of Medan. With a throaty full-bodied laugh, he describes their playing as terrible, bordering on disastrous. They had never played together before that day, and were versed in completely different village styles. Thus the patterns they knew were not *cocok*, suited, to one another. Instead of flawlessly interlocking, they were constantly crashing together, *kom* (o) and *peng* (e) or *Dag* (D) and *Tut* (T) not deftly dancing around one another but colliding. To keep from ruining the whole performance, Tama settled on just one pattern that could work with anything Gobleg played, and only rarely varied it at a surface level. This despite the fact that both were expert improvisers in the same genre, who,

had they been playing with their regular partners, could have created *cocok* musical interactions for hours with very little effort.

On a practical, concrete level, *pasangan* or partnership takes the form of specific pre-composed motives that regular *kendang arja* partners create together, sometimes called *rumus* (see [11: 79–81]). When improvising in an 8-beat cyclic structure, if Pak Dewa were to play beats 4 and 5 using the grey *wadon* gesture in Fig. 45, Cok Alit, and no other drummer, would know to respond, in beats 6 and 7, with the black *lanang* gesture. This call-and-response is a *rumus* pattern that the two have developed together and shared with their students. Were Pak Dewa to play an *arja* performance with Pak Tut, for instance, the latter would not know how to respond to this series of *Dag* (D) in the appropriate manner.

Yet partnership goes far beyond *rumus* and becomes a shared knowledge base: a mutual understanding of the principles of *arja* interlocking. Cok Alit, having played for many years with Pak Dewa, knows which patterns will work well with his partner's collection of patterns, and how to vary each one appropriately in performance. What's more, he has developed new patterns over the years with that specific interaction in mind. Of course the complexity of both drummers' improvisations means that collisions do happen, even the more egregious collisions of *Dag* (D) and *Tut* (T). To illustrate the unpredictability of interlocking in the course of performance, Fig. 46 shows the final four beats of Pak Dewa's most basic *wadon* pattern, *dasar*, with several improvised pairings from a recording session with Cok Alit. Inherent in the *wadon* pattern itself is a potential collision on the *Dag* (D) stroke directly preceding beat 2. Because the most common placement for a *Tut* (T) stroke in his partner's idiom, as we've seen, is the 4th subdivision of a beat, a *Dag-Tut* collision is a distinct possibility anytime Pak Dewa plays this pattern or one of its variants. In fully pre-composed interlocking, the *lanang* pattern would be made to work around this *Dag* (D), as it does in the first example in Fig. 46. Yet a *Dag-Tut* collision in this location can be acceptable in *kendang arja*, provided the more important signposting-*ngegong*in *Dag* (D) on beat 3 sounds without a *Tut* (T) collision. That said, some of the pairings in Fig. 46 do contain a less desirable level of collision, shown in the composite patterns with dashed rectangles.

Each of the improvised interactions in Fig. 46 contains surface-level variation that will not be addressed. What is relevant here are the differing levels of *cocok*-ness, suitability, in each pairing. While the drummers' shared history means that none of these levels of collision is disastrous, the coincidence of each of these pairs

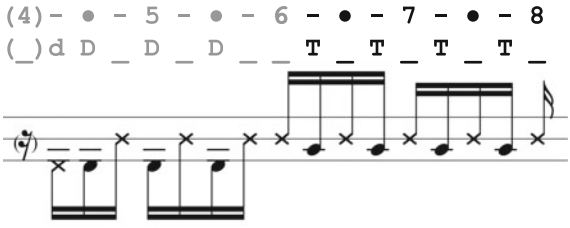


Fig. 45 Partner patterns: *rumus*

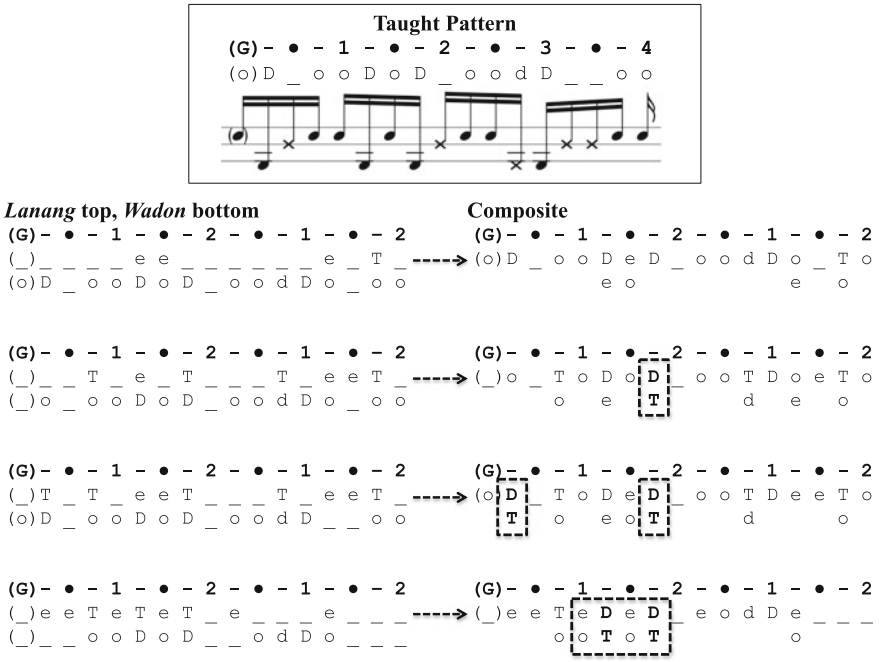


Fig. 46 Chance interactions in improvised performance

of patterns in improvised performance, the more and the less *cocok*, speaks to the controlled arbitrariness of *kendang arja* interlocking. Yet looking back at the *Dag-Tut* interactions in Fig. 7, we can see that, even with this unpredictability, such collisions are rare among seasoned partners.

Pasangan, partnership, guides all aspects of improvisation: each drummer must become accustomed to his partner's gestural communications and rhythmic cues as well as his grammar of patterns and variations. As we saw in Fig. 1 and Video 1, longtime *arja* partners sit close together while playing, reading cues from subtle movements and facial expressions caught out of the corner of the eye as often as sound: a communication style perfected through years of mutual musical experience. This depth of connection between partners is central to their successful *arja* improvising. As we have seen, the spoken and unspoken guidelines for *kendang arja* are general and flexible enough that specific interpretations are delightfully varied, and sometimes incompatible. Yet within each village or individual style, though there may be a near infinite corpus of patterns at play, is a largely predictable palette of pattern types and idiomatic methods for varying them. A drummer can learn to fit perfectly into her partner's musical grammar through friendship, time, dedication, practice: *pasangan*.

8 Concluding Thoughts

One of the exciting things about being a fieldworker as well as an analyst is that discoveries in one modality can inform and give focus to the other, creating a feedback loop of inspiration and insight. The analyses I have made of *kendang arja* patterns from different drummers throughout Bali would have been impossible without personal connections and an ethnographic approach. Only through attempting to perform with these musicians, learning directly from them, and spending hours in their homes discussing music did I slowly unearth each one's personal theories on the roles of each drum and their interactions with one another: the relative simplicity of the *lanang*'s improvisations, the structural importance of *Dag* (D) and *Tut* (T), the on-beat-off-beat guidelines, the various applications of *ngegongin*, and the segmentability and rearrangeability of pattern elements. These concepts then became a doorway to unraveling the unspoken strategies of *arja* performance practice: a guide to the analysis of improvisation.

8.1 Ethnographically-Informed Analysis and Computational Archives

This chapter's analyses were originally presented at a symposium on computational archiving. At first glance this is a strange fit, perhaps. Yet many ethnomusicologists have large collections of field recordings. Important resources in and of themselves, if framed in a more holistic way, these could provide fodder for collaborative research and cross-genre analysis. To become a powerful research tool rather than a simple storage facility, an ethnomusicological sound archive needs contextualizing. I imagine a public repository of *kendang arja* recordings carefully categorized by village, performer, and cyclic structure, each track contextualized with information on the musicians, their pedagogical approaches and oral music theories, and my interpretations of these. Such a collection might include full transcriptions of improvised sessions and, importantly, would feature isolated transcriptions of each drummer's taught patterns. With this more complete information package, researchers from multiple disciplines could then approach the collection from a place of knowledge.

In 2016, in collaboration with my MIT music colleague Michael Cuthbert, I began work with an undergraduate student in computer science, Katherine Young, to develop genre-appropriate computational analyses of *kendang arja*. Such an approach was far outside my skill-sets. But with access to my transcriptions and my knowledge about the music's cultural and music-theory contexts, Katherine was able to build on this research in new ways. I see broad potential for a digital sound archive to foster this kind of collaborative and multi-author work across institutions. For *kendang arja* alone, those involved in digital transcription could work on recordings I've yet to transcribe; those interested in computational analysis could build on Katherine's work; those wishing to analyse micro-timing or acoustics or music infor-

mation retrieval in *arja* could do so from a more culturally-informed perspective. Were there a user-generated platform linked to the archive, other ethnomusicologists studying contrasting styles of *arja* could post their findings: recordings, transcriptions, taught patterns, musician profiles and lineages, and oral music theories, as well as the researcher's own theories, analyses, and comparative observations. All this would enable an intertextual approach to *arja* research, where scholars from diverse institutions could access one another's recordings, working collaboratively toward a larger theory of *arja*. And if such annotated collections existed for numerous practices worldwide, we could finally engage in a true comparative, cross-cultural, and cross-disciplinary analysis of improvisation. The computational archive and the analytical ethnomusicologist could then develop a valuable and unexpected new *pasangan*.

References

1. Asnawa IKG (1991) The *Kendang Gambuh* in Balinese music. M.A. thesis, University of Maryland, Baltimore County
2. Berliner PF (1994) Thinking in Jazz: the infinite art of improvisation. University of Chicago Press, Chicago
3. Brinner B (1995) Knowing music, making music: Javanese *gamelan* and the theory of musical competence and interaction. University of Chicago Press, Chicago
4. Charry E (2000) Mande music: traditional and modern music of the Maninka and Mandinka of Western Africa. University of Chicago Press, Chicago
5. Chernoff JM (1979) African rhythm and African sensibility: aesthetics and social action in African music idioms. University of Chicago Press, Chicago
6. Cook N (2013) Beyond the score: music as performance. Oxford University Press, Oxford
7. Cook N, Everist M (eds) (1999) Rethinking music. Oxford University Press, Oxford
8. Dibia W (1992) *Arja*: a sung dance-drama of Bali: a study of change and transformation. Ph.D. dissertation, University of California at Los Angeles
9. Ellingson T (1992) Transcription. In: Myers H (ed) Ethnomusicology: an introduction. MacMillan, London
10. Hall ET (1992) Improvisation as an acquired, multilevel process. Ethnomusicology 36(2):223–235
11. Hood MM (2001) The kendang arja: improvised paired drumming in Balinese music. M.A. thesis, University of Hawai'i
12. Hood MM (2002) Improvised paired drumming: the ngematin/ngegongin relationship of kendang arja. Bheri 1(1):71–83
13. Locke D (1998) Drum Gahu: an introduction to African rhythm. White Cliffs Media Co, Crown Point
14. Marian-Bălaşa M (ed) (2005) Notation, transcription, visual representation. World Music (special issue) 47(2)
15. Mapaya MG (2014) Dinaka/kiba: a descriptive analysis of a northern sotho song-dance performative compound. Afr J Phys Health Educ Recreat Dance (AJPHERD) 20(2):426–438
16. Monson I (1996) Saying something: jazz improvisation and interaction. University of Chicago Press, Chicago
17. Redhead L, Vanessa H (2016). Music and/as process. Cambridge Scholars Publishing, Newcastle upon Tyne
18. Stanyek J (ed) (2014) Forum on transcription. Twent Century Music 11:101–161

19. Tenzer M (2000) Gamelan Gong Kebyar: the art of twentieth-century Balinese music. University of Chicago Press, Chicago
20. Tilley L (2013). *Kendang Arja:* the transmission, diffusion, and transformation(s) of an improvised Balinese drumming style. Ph.D. dissertation, University of British Columbia
21. Tilley L (2014) Dialect, diffusion, and Balinese drumming: using sociolinguistic models for the analysis of regional variation in *Kendang Arja*. Ethnomusicology 58(3):481–505
22. Weinberg N (1994) Guidelines for drumset notation. Percussive Notes 32(3):15–26
23. Widjaja NLNS (2007) Dramatari *Gambuh* dan Pengaruhnya pada Dramatari Opera *Arja*. Ph.D. dissertation, Universitas Gadja Mada

Temperament in Tuning Systems of Southeast Asia and Ancient India

Rolf Bader

Abstract Tuning systems in many musical cultures in Southeast Asia as well as in India are often considerably different from Western tuning systems. Within these cultures, the literature as well as oral tradition on theoretical reflections of why choosing which tuning system are scarce. For Western scholars, only considering the pitches played or the tuning of instruments is not perfectly satisfying, as it lacks insight into the reasons for certain tunings. Still when considering ecological constraints, like ensemble setups, instrument building, or the acoustics involved, often the reasons for choosing a special tuning becomes clear. The paper presents examples from fieldwork in this regions between 1999 and 2014, and often finds musicians wanting to get close to just intonation, but deviating from it mainly because of three constraints: The need to switch between different ensemble types, constraints of instrument building, and acoustical constraints. The paper therefore suggests that many tunings in Southeast Asia are temperaments in the sense of compromises between a desired system, often just intonation, and such constraints.

1 Introduction

Tuning systems in Southeast Asia as well as in many other non-Western ethnic groups have often been found not to be built of just intervals. These tunings also deviate strongly from a Western twelve-tone equal tuning temperament. Ellis proposed the scale of Siam, todays Thailand, to be equidistant [1, 2], a scale also associated with the ancient kingdom of Ayutthaya [3]. Here each of the seven tone steps is 1200 cent/7 = 171.4 cent. Such a system is built in a simple way. Still contrary to the Western tradition there is no known background in the music theory of this tradition

Electronic supplementary material The online version of this chapter (https://doi.org/10.1007/978-3-030-02695-0_3) contains supplementary material, which is available to authorized users.

R. Bader (✉)
Institute of Systematic Musicology, University of Hamburg, Hamburg, Germany
e-mail: R_Bader@t-online.de

R. Bader (ed.), *Computational Phonogram Archiving*, Current Research in Systematic Musicology 5, https://doi.org/10.1007/978-3-030-02695-0_3

in why such an equidistant scale is used. In reality, these tunings may also deviate a lot from the theoretical intervals [4]. Tunings in Cambodia or Myanmar may differ highly from the equidistant scale and also differ between musical styles. Indonesian *gamelan* tunings even differ between ensembles and the reasons for instrument builders choosing a certain tuning are still under debate [5–7]. It appears that the choice of a tuning system is not arbitrary for builders and musicians alike. Still the reason for choosing just the present scales is not clear yet.

The main problem when investigating non-Western tuning systems is the scarce literature on music theory discussing tuning systems as well as a lack of reflection among instrument builders and musicians on this topic, in strong contrast to the huge amount of Western literature on the subject (see [1, 5–8] as examples). Therefore only measurements of the actual intonation in the performance context can give insight into the underlying system. Measurements display tuning systems which seem quite arbitrary at first and fitting a theory to them is not straightforward. This may lead to the conclusion that the systems are only based on habits and customs and do not have any theoretical background at all. Many authors in their fieldwork find musicians not to be aware of any music theory. This might seem supported by the general absence of music theory in musicians and instrument builders beyond simple features like solmization, scale, phrase names or basic tunings of the instruments. Still such systems may be present, either built by constraints of any kind or by previous musicians or instrument builders who incorporated their knowledge and systems into the building process or intonation into musical instrument geometries or music education.

Indeed, considering these tuning systems as random might be wrong. From acoustical, physiological and psychological standpoints, the just-tone system is likely to be preferred in terms of the absence of roughness when two notes with harmonic overtone series are played simultaneously [9]. However from the standpoint of music instrument building, the construction of instruments with a just-tone tuning is not at all trivial [10, 11]. Wind instruments have complex bore profiles and horns, which prolong the length of the sounding tube, hence adding an end-correction. As this end-correction depends on frequency, the pitches played are never in a perfect harmonic relation. Musicians therefore need to intonate the instruments to adjust the pitches by changing lung pressure, lip tension or the distance from the lips to a labium. Tuning systems of ancient musical instruments have often been derived from their finger hole placements alone which cannot result in the pitches actually played due to all these articulation parameters. Interestingly, in Southeast Asia many flutes and shawms have an equidistant finger hole placement. However such a placement does not lead either to an equidistant scale or to a just intonation. As musicians have strong influence on the pitch played with their playing technique, such wind instruments may be played in an equidistant or a just intonation as will be discussed in this paper for music of the Kachin in Myanmar and Singhalese *puja* temple music.

The acoustical problem of different tube length for single frequencies leads to another interesting finding. The inharmonic overtone structure present in all wind instruments might be expected to result in an inharmonic sound when playing a single note. However this is not the case and when playing with normal pressure

articulation, the resulting tones show a very precise harmonic series. This is caused by the highly complex nature of the driving system of these instruments leading to a synchronization of the overtones into a strictly harmonic series [12]. This happens with reed and double reed instruments, labium instruments such as flutes or flue organ pipes, as well as with the singing voice. Harmonic overtone series are indeed preferred in the West, which is strikingly shown by the fact that many Western percussion instruments, like bells or xylophones are built with such geometries to meet as many harmonic overtones as possible. Still from a physical standpoint of voice and of musical instruments tone production, this is not at all straightforward. Evolution has produced highly complex physiological systems producing such harmonic series, most likely because they allow information transport with very low energy supply. As such complex systems are present mainly with human and animal sound production, hearing a harmonic sound will let a listener assume the presence of a living being with a certain intelligence and will, which needs to catch our attention. Therefore, building musical instruments with harmonic series is building instruments with a basic semantic and intelligence property.

Harmonic sounds are fundamentally different from inharmonic sounds or from random noise. A just-tone system has the same mathematical relations as a harmonic overtone series. Although tuning systems do not directly emanate from overtone structures, just intonation still references the harmonic spectrum, in which the octave (2:1), fifth (3:2), fourth (4:3), major (5:4) and minor third (6:5) and even the Pythagorean major second (9:8) are present. Nevertheless, both harmonic and inharmonic sounds are referring only to the harmonic overtone series simultaneously present in time and not to a tuning system played as a melody at adjacent points in time. Only in polyphonic ensembles are two or more sounds played at the same time and indeed inharmonicity catches our attention much stronger there.

Many musical instruments in Southeast Asia are percussion instruments with no such exact harmonic overtone structure. This does not mean that their spectra are arbitrary. Drums have a complex spectrum where some strong overtones build a nearly perfect harmonic series allowing melodies to be played as discussed with the *bama pat waing* drum circle below. Gamelan instruments, such as the *saron* or the *gender*, have bamboo tubes resonating mainly at the fundamental frequency enhancing pitch perception at this frequency. The Helmholtz reasoning of a tuning system emanating from the least rough intervals has also been enlarged to inharmonic sound with some interesting results discussed below in more detail [5]. Still such roughness might be musically interesting and implemented by instrument builders like with the *ombak* of gamelan gongs which are casted not perfectly round producing degenerated modes with slightly different frequencies resulting in amplitude beatings. Also Indonesian *gender* metalophones such as that of the *gamelan wayang* have such beatings, where instruments are paired and two bars with the same pitch are tuned 2–4 Hz apart [13]. It has been suggested that these pitch deviations are the reason for stretched octaves in *gong kebyar* [7]. In this case, the tuning of stretched octaves is not random but based on a system of acoustic and physiological constraints.

So considering acoustical, physiological and psychological standpoints, tuning systems are very likely not arbitrary. This leads to the discussion of musical universals

which have been proposed by comparative and systematic musicology following the Hornbostel/Stumpf/Sachs tradition. This tradition has led to reasonable results such as the Hornbostel-Sachs instrument classification system but did also fail as with the theory of blown fifth. Recent investigations of statistical nature involving large ethnomusicological recordings, such as the cantometrics of Alan Lomax, or more recent investigations using the Garland encyclopedia recordings [14] do find features reappearing more or less often in music. This approach tells hard universals, features always present everywhere from features present only within statistical likelihood. Although these data are often surprising, like the rareness of pentatonic scales, this approach does not give deep insight into the reasons for the statistical presence or absence of musical features. Structural approaches might give more insight here like musical meaning in South Indian temple *pujas* music relating the structure of the ritual with that of the rhythms and melodies [15]. Obviously the acoustical examples discussed above also support a structural or systematic explanation.

Such structural systems often explain deviations from just systems by combining a set of acoustical, physiological, psychological, ecological and other constraints. In terms of tuning systems, this is very well known in the West and called temperament in the original sense of 'to temper' or 'to make a compromise'. The Western temperaments suggested that the mean-tone tuning of Renaissance music, which uses many just third intervals, or the mainly fifth-oriented systems distributing the Pythagorean comma to a certain amount of fifth, like Werckmeister, Kirnberger, Valotti or Young, all try to temper roughness of intervals to enlarge the amount of playable intervals. Here we have a just intonation as a desired system and a detuning according to a constraint, which is the amount of playable tonalities. Still this is not the only temperament used in Western music. A modern piano is tuned with stretched octaves which lead to an overall 's-shape' of temperament of the instrument [16], with the bass registers tuned flat and the treble registers tuned sharp. As a consequence, all intervals are stretched, e.g. the octave is more than 1200 cent. Modern keyboard music uses the twelve-tone equal temperament where each tone step is 100 cents which is a distribution of the Pythagorean comma to all twelve fifth. When implementing this to the guitar using logarithmically spaced frets, this does not lead to a tuning system of 100-cent intervals due to the different string tensions. When pressing all six strings down at one fret, the finger pressure necessary to do so varies between the strings leading to a different detuning at each tone. Automatic guitar tuning machines, having implemented many different tunings, also have a 'guitar tuning' taking this mistuning into consideration. Therefore guitars are never in tune, but Western listeners are so accustomed to this detuning that they accept it as 'in tune'. Here the compromise or temperament is caused by a physical or material constraint.

In Western music, these tunings are rationalized in musical acoustics as well as in music theory, making it accessible to non-Westerners. Without explaining texts, a non-Western researcher approaching Western tuning systems might find it equally hard to understand the wide variety of measured tunings as Westerners find it hard to understand the tuning systems of Southeast Asia. Indeed in cases where there are theoretical reflections of tunings systems in Southeast Asia, the situation is much

clearer and tempering appears also caused by historical interactions of ethnic groups [3, 17, 18], suggesting mixing and compromising between musical tuning systems.

Temperament is another word for compromise in Western tuning systems. Why should this not also be the case for tuning systems in Southeast Asia? Alike Western examples, the constraints of Southeast Asian music tempering might be the need to play different tonalities for different tunes, or they might be constraints of instrument building or the need to play with changing ensembles of different musical styles, may they be within the culture or such of other ethnic groups. Following this idea, universals might be defined as a system of constraints leading to a compromise, a temperament, a tuning system.

As discussed above, a pure statistical approach to universals is lacking of many constraints leading to the present tonal system. Still this approach has the advantage of a large data collection. Another approach is to discuss single examples found during ethnomusicological fieldwork where the context can be taken into consideration. For example, Vetter found the tuning of a *gamelan* as a compromise between javanese *sléndro* and *pélog* scales and musical phrases played as well as vocal constraints as he witnessed a gamelan tuning by a professional tuner [6]. Of course, this approach cannot discuss a large database which is a trade-off. This paper tries this way by extracting some constraints leading to musical temperaments found in the measured tunings.

The material is based on fieldwork done by the author between 1999 and 2014 in Cambodia, Thailand, Myanmar, Sri Lanka, Bali and India, recording traditional music, collecting musical instruments and analyzing the tuning systems, instrument geometries and comments of musicians and instrument builders. It will argue that even complex tuning systems found with the instruments can be explained by compromises between simple tuning systems, instrument geometry or playing conditions as well as instrument building processes.

In a first section, some discussions about tunings are presented, starting from ancient Indian music theory. Due to the rareness of music theory in Southeast Asia, the ancient Indian system is discussed as an example of a temperament using equidistant intervals. Whether this system is the basis of other equidistant examples, such as the heptatonic equidistant Thai scale or the geometrically equidistant finger hole placing of many wind instruments, cannot be decided on historical grounds, although it might be the case from the content of the later finding. Then selected instrument measurements are discussed, both as recorded in the field and as investigated at the laboratory of the Institute of Systematic Musicology in Hamburg where they could be explored in depth. Along with the measurements, the context of the music as well as the possible constraints found for the tuning system are discussed. In a third section, findings are summarized and a list of constraints leading to the present temperaments is discussed.

2 Equidistant Tunings

Ancient music theories of India and Sri Lanka differ from modern *raga svara* (tone step) and *that* (scale) in their notion of a *śruti*. Today Indian music theory knows twelve tone steps, similar to Western scale steps, and uses *śrutis* only as melismatic deviations from the mostly heptatonic *gama* (tone row), a selection of notes from the twelve main steps [19]. This is different when examining ancient music theories, the most prominent being the *Natyashastra* of Bharata written probably 200 BC–200 AC. The main source of ancient Tamil music theory is the epic *Cilappatikāram*, studied extensively by [20]. Yet another music theory, the *Saṅgītopaniṣat-sāroddhāraḥ,* is special as it associates the tones with attributes of deities, colors, etc. and also discusses melisma in detail which are therefore described as tantric [21]. It was probably written in West India where influence of Western traders was strong and therefore a more liberal society may have accepted such a deviation from the mainstream music theory. As it was written probably in the 14th century AC, we may conclude that the ancient theories continued well into the Moghul period of Indian history.

The difference in all three music theories compared to modern Indian music theory is that the *śrutis* divide the octave into 22 equal intervals which means each *śruti* is 54.5 cents. As Rowell points out, the term *śruti* in ancient music philosophy found in the *raknasangita* has two meanings. Literally it means 'what is heard' and, in text interpretations, is the opposite to *smriti*, what is remembered. So *śruti* points to something eternal or theoretical rather than to something present. This meets very well the use in music theory, as a scale played seldom consists of 22 tone steps but mostly of seven, where the *śruti* are only the backbone of possible steps. Secondly, *śruti* in the old texts corresponds to just-noticeable differences (JND) which could point to the small interval of a *śruti* of 54.5 cents. This does not correspond to the hearing threshold of pitches, which is much smaller and very much frequency dependent. In musical situations, a difference of a semitone is practically always heard, a quarter-or quaver-tone step is most often perceived, and smaller intervals play a role producing beatings which is extensively used in gamelan music, where the *ombak* of a *gong gede* or the tuning of two gender in ga*melan wayang* (for puppet theater) in Bali tunes the paired instruments 2 or 4 Hz apart. Therefore *śruti* as JND makes sense in a solo performance, where no beating can occur and where tuning is important but not in a very strict sense.

When comparing the tunings of the different scale steps of the three music theories to the just intonation, it clearly appears that many scale steps come very close to just temperament. Thirteen *śrutis* arrive at 708 cents which is close to the 702 cents of a 3:2 just fifth interval, nine *śrutis* with 490 cents are also close to 498 cents of a 4:3 just fourth. Also seven *śrutis* with 381 cents are close to the 5:4 just major third with 386 cents. This continues with four *śrutis* at 218 cents which are not too far from 204 cents of just major second of 9:8, 109 cents of two *śrutsi* come close to 112 cents of a just minor second of 16:15, 15 *śrutis* with 818 cents are close to 814 cents of a just minor sixth of 8:5, the major sixth 5:3 with 884 cents has its nearest neighbor at 16 *śrutis* at 872 cents, or twenty *śrutis* with 1090 cents are closest to the major

Table 1 Comparison of *śrutis* and *māttirai* taken from Tamil epic *Cilappatikāram* (200–500 AC) as reconstructed by [20], *Natyashastra* (*200 BC–200 AC*) and the tantric music theory *Saṅgītopaniṣat-sāroddhāraḥ* (1350 AC), the latter two in their fundamental modes only. All divide the octave into 22 *śrutis* of equal size and find scale steps by intervals of two, three or four *śrutis*. When comparing the tuning with just intonation (for major and minor second, third, sixth, seventh, the fourth, triton and fifth), many intervals of the ancient music theories come very close to a just scale

Tamil ***Cilappatikāram***	*kural*	*māttirai*	*tuttam*	*māttirai*	*kaikkila*	*māttirai*	*Ulai*	*Māttirai*	*Ili*	*māttirai*	*vilari*	*māttirai*	*taram*
pālaiyāḻ		4		3		2		4		3		2	
Mixolydisch	0		218		381		490		708		818		1035
kuṟiñchiyāḻ		4		3		2		4		3		4	
Ionic	0		218		381		490		708		818		1090
marutayāḻ		4		3		4		2		3		4	
Lydisch	0		218		381		600		708		872		1090
neytaliyāḻ		2		3		4		2		3		4	
Lokrisch	0		109		272		490		600		763		981
cevvaḻiyāḻ		2		3		4		4		2		3	
Phrygisch	0		109		272		490		708		818		981
Nadyashastra	*Sa*	*śruti*	*Re*	*śruti*	*Ga*	*śruti*	*Ma*	*śruti*	*Pa*	*śruti*	*Dha*	*śruti*	*Ni*
Śrutis		4		3		2		4		4		3	
	0		218		381		490		708		926		1090
Saṅgītopaniṣat-sāroddhāraḥ	*Sa*		*Re*		*Ga*		*Ma*		*Pa*		*Dha*		*Ni*
Śrutis		4		3		2		4		3		4	
	0		218		381		490		708		872		1090
Just intonation	**0**		**204** **112**		**386** **316**		**498** **610**		**702**		**884** **814**		**1088** **1018**

seventh 15:8 with 1088 cents. Only the minor seventh 9:5 with 1018 cents does find 981 cents of 18 *śrutis* as its nearest neighbor which is more far apart. An interesting exception is the minor third 6:5 with 316 cents. The closest interval in Table 1 is five *śrutis* with 272 cents which is 44 cents apart. Using six *śrutis*, we would arrive at 327 cents which is only eleven cents from the just minor third. It seems that the just minor third was not wanted in the scales.

We find an equal tuning of 22 intervals in the octave to be a very good compromise when approaching just intonation. Comparing this *śruti* compromise with the Western 12-tone equal tuning of 100 cents per step, by summing all deviations of all intervals (without the tritone and using the better minor third with the *śrutis*), we have 102 cents difference for the *śrutis* case and 109 cents difference for the Western equal tuning. Although this is not considerable, a 22-tone case is of course more flexible than a twelve-tone one. On the other hand, with keyboard instruments for example, we would need 22 keys per octave and not only 12, which is not practical.

Equidistant tuning was also reported in Western traditional music. Pondus [22] made extensive studies in bagpipe music in several European countries. He finds only the fifth and the second scale step to be almost in accordance with just tones while the tuning of the other steps show considerable deviations. He also found seven-tone

Table 2 Theoretical tunings of the Bama in Myanmar with neural intervals compared to the Ayutthaya heptatonic equal tuning. Data from Muriel Williams 2000

Tuning system	Step 0	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	Step 7
Hyin-lon	0	200	350	500	700	900	1050	1200
Auk-pyan	0	200	350	550	700	900	1050	1200
Pale	0	150	350	550	650	850	1050	1200
Myin-zaing	0	150	350	550	700	850	1050	1200
Ayutthaya	0	171	342	514	685	857	1028	1200

equidistant tunings of Italian pipes from the 15th century and, in the 70th of the 20th century, still some examples of pipers in Miranda do Douro in northeastern Portugal who tune their instruments according to an equidistant temperament.

Of course, in the West, there is the twelve-tone equal tuning used today in Western electronic music, which is an equidistant tuning, too. This tuning is not found with pianos which are tuned with stretched octaves and which increase towards high pitches and decrease towards the lower end of piano keys [16]. The guitar and electric fretted bass both show equidistant tuning in their fretting. Still because of string thickness and differing finger pressure needed to play the strings, their actual tuning deviates from equidistant tuning a lot. The deviations in the piano tuning are intended, that of the guitar tuning is accepted, but both are only theoretically based on equidistant tuning and in reality we have been accustomed to these tunings. Wind instruments are very hard to tune because of their bore diameter, cup flanging and driving mechanism and players need to adjust pitches while playing by changing playing pressure or the distance between mouth and mouthpiece. Many pitches played are thus compromises between the real sound and the desired pitch, may this be just, equidistant or according to a historical tuning, such as mean tone or else. A truly equidistant sounding scale is found nearly solely with electronic music.

3 Thai and Cambodian Systems

The tuning system of Thailand is very much established to be a heptatonic equidistant scale where each scale step has 171.429 cents, incorporated in Table 2 as Ayutthaya scale. As discussed by Williamson the Bama, after defeating Ayutthaya in 1767 took over the Thai scale to some extend [3]. Therefore this Ayutthaya scale seems to be at least of such an age.

In Cambodia, the Red Khmer, during their reign of 1975–1978 and until the UNU organized elections in 1995 in the north of Cambodia, are known for killing intellectuals, musicians and monks and for abandoning music. Although many forced marriages of monks took place and monasteries were destroyed, there seems to be no evidence for a systematic destruction of monasteries [23]. In terms of *smot* Buddhist chanting discussed below, Francois Bizot, who was arrested and investigated by the

Khmer Rouge, reports that *smot* singing was allowed and also performed by Red Khmers with slightly different lyrics, where the Buddhist community *sangha* was substituted by the word for the Communist party *ankar* [24]. Also *mohori* wedding and entertainment music, where the equidistant tuned *takhe* string instrument is played as discussed below, was allowed to some extend for entertainment as reported by Svam Prum, who was living under Red Khmer dictatorship as a child.[1]

Sam [25] notes that there is no word for scale in Cambodian language and no written sources on this issue. He reports of two fundamental scales, a pentatonic and a heptatonic one, where he finds the precise intonation of the pitches in the notes played and a central tone near a Western G of the *pinn peat* ensemble referring to a *roneat ek* xylophone key, a *khloy* flute fingering or a stringed *thake* fret, without going into the details in terms of tuning systems.

4 Myanmar Bama Tunings

The traditional tuning of Bama music is the *hyin-lon* scale [3] with a neural 3rd and a neutral 7th as shown in Table 2. Williamson reports that after the Bama had defeated Ayutthaya in 1767, the Thai music with the Ramayana plays became very popular among the Bama. As the Thai heptatonic equidistant scale deviated from the *hyin-lon* scale, compromise scales *auk-pyan* and *pale* were developed. These scales deviated so much from *hyin-lon* that the musicians deeded to retune their instruments. As another compromise between these scales, a new scale *myin-zaing* was developed but not very much used. This scale should meet both the old *hyin-lon* and the newer scales and therefore would not need a retuning of the *saun gauk* Burmese harp or *pat wain* drum circle tuning.

5 Indonesian Tuning Systems

Tenzer discusses the origin of Balinese g*ong kebyar* as a reaction to the late colonization of Bali (1904 and 1908) where many of the upper class committed suicide in front of the colonizing soldiers [7]. In response, reading contests have been performed using the old Hindu epics Ramayana and Mahabharata accompanied by *gamelan*. As these contests were about the speed of reading, the music followed building the fast *gong kebyar* style. When building new bronze instruments, the tuning systems seem to have changed over the twentieth century which prefer to narrow the small intervals even more, down to about 50 cents, thus enlarging the larger tone steps up to about a major third [7]. Using the Balinese solmization *ding, dong, deng, dung, dang*, the older *tirus* tuning has 104 cents between *dung* and *dang* as a minimum and 372 cents between *dang* and *ding* as the maximum interval. The modern *begbeg*

[1] Svam Pruhm, personal communication.

tuning used for *gong kebyar* has 81 cents between *dung* and *dang* and therefore a much wider 453 cents interval between *dang* and *ding*. A middle *sedeng* tuning in between *begbeg* and *tirus* is also known. In addition, Tenzer discusses the enlarged octave in *gong kebyar* tuning of over 1200 cents and relates it the detuning of pairs of metalophones, where one instrument is tuned 6.3–9.7 Hz off. The detuning results in a characteristic beating called *ombak* with beating speed called *penjorog*. The lower instrument is called *pengumbang* and the higher, *pengisep*. While one instrument of the pair would always be tuned to perfect 1200 cents octaves, the other would always be too low or too high. Therefore the octaves are necessarily deviating from a perfect octave, where measurements show that lower and higher registers are stretched while the medium range is flatter. This points to an impact on the timbre of the tuning system.

6 Timbre-Based Tonal System Theory

Helmholtz: *Tonal Systems as Minimization of Roughness*

Perceptual roughness is proposed by Helmholtz as the foundation of a tonal system based on simple integer intervals like 2:1 of an octave, 3:2 of a fifth, 4:3 of a fourth, 5:4 of a minor third, etc. He argues that two frequencies f1 and f2 close together have a beating, an amplitude modulation of a periodicity fdiff = f2 – f1. When fdiff increases, the beating becomes rough with a maximum roughness at fdiff = 33 Hz and decreases again with higher fdiff. When two notes which both have a harmonic overtone structure are played, some partials from both tones may be close together and their difference fdiff will determine the amount of roughness. As this might happen with several combinations of partials, the roughness may add up to a certain amount. Helmholtz systematically calculates this roughness for all tone distances from unison 1:1 to the octave 2:1. At integer ratios, the roughness curve has minima and hence the tones played are least rough. He then concludes that we prefer the simple integer ratios for tonal intervals not because of their simple ratios but because of the timbre or sound two tones have when played together, the sound of least roughness. Therefore Helmholtz is the first to derive tonal systems not from numbers and ratios alone but as well from the overtone structure of pitches and tones, from timbre.

Sethares: *Minimization of Roughness as Basis of Javanese Gamelan Tonal Systems*

Sethares enlarges this point of view to Indonesian gamelan instruments. He investigates two Javanese *gamelan* in both *pelog* and *sléndro* tunings, *Swastigita* and *Kyai Kaduk Manis*. Measuring the keys of several *gender, saron* and *boning*, he calculates mean tunings for the ensembles, where indeed *sléndro* comes very close to a equidistant pentatonic scale of tone steps in a range of 234–248 Hz. The *pelog* tuning is complex as expected with tone steps 119, 155, 282, 119, 108, 197, and 244 cents for the *Swastigitha* and 157, 153, 266, 110, 126, 183, and 217 cents for the *Kyai Kaduk Manis* ensemble. For his reasoning, he uses two instruments from the ensemble, the

bonang with the *pelog* and the *saron* for the *sléndro* tuning. As he considers the timbre of the instruments, he first needs to extract the inharmonic overtones of the instruments. He finds relative frequencies to the fundamental as

Bonang Swastigitha:	1	1.52	3.45	3.92		
saron Swastigitha:	1	2.76	4.72	5.92		
saron Kyai Kaduk Manis:	1	2.39	2.78	4.75	5.08	5.96

Sethares then uses a similar method to Helmholtz's, which is to calculate the roughness of these sounds by comparing the relative frequencies of partials. Therefore he needs two tones where the first is one of the spectra shown above and the other is a harmonic overtone series of 1:2:3 … Taking such a harmonic series rather than the spectra of the other instruments the *bonang* or *saron* are playing is not straightforward. Gamelan builders tune the instrument during the building process using a *suling* bamboo flute with a harmonic overtone series. A flute has only weak harmonics, and when the gamelan is playing, no *suling* or other instrument with a harmonic overtone series are playing along, sometimes with the exception of the *rebab*, a bowed fiddle. The roughness curves correspond well with the mean tuning of the sets as calculated from the measured instruments. This is indeed a remarkable finding but does hold for the used overtones compared to a harmonic series which is not present in the gamelan set as discussed. Furthermore a *saron* of a *Javanese gamelan* is dominated by its fundamental frequency and does not have many strong higher harmonics able to produce roughness. Still the basic idea is interesting as it points to a universal musical tuning; the minimization of perceptual roughness also for the music of Southeast Asia.

The roughness minimization may play a larger role with Balinese *gong kebyar* as this music is much brighter in sound with much stronger higher partials able to add serious roughness [13]. Figure 1 shows the *gender dasa* used for the investigation (top). The plates have a trapezoid shape (middle) which is schematically shown in the bottom plot. This shape is more difficult to craft than a flat shape and thus it is interesting to see if there is any influence of this shape on the sound of the instrument. Using a Finite-Difference Time Domain method, the sound of the *gender* plate is calculated by virtually striking it with a hammer. As displayed in Fig. 1, the flat geometry sound shows much less higher harmonics (top) as the trapezoid one (bottom). Thus the Balinese *gender* plate shape is built to enhance many higher harmonics by this trapezoid shape. Although the reasons are not perfectly clear, the kinks of the plate seem to produce higher harmonics by diffraction of lower frequencies. This process may be similar to the brightness of the Chinese *tam-tam* or Turkey crash cymbals which produce their additional higher harmonics by the hammering of the metal. These higher partials last long enough to participate in roughness production.

7 Cambodian *Thake*

The Cambodian *thake* is a stringed fretted instrument with three strings played by a wooden stick. It is used mainly in *mahori* bands performing at weddings and secular feasts and also today as a tourist attraction. The frets are glued to the body and therefore the tuning (*go ma* in *khmer*, lit. 'child', therefore denoting something small) is fixed by the instrument builder. When visually examining the instrument, the frets clearly appear in a logarithmic relation similar to that of a Western guitar fretboard. Examining the octave relations, only seven tones within one octave are present, not the twelve tones of guitars. Hence the tuning strongly points to a seven-tone equal temperament.

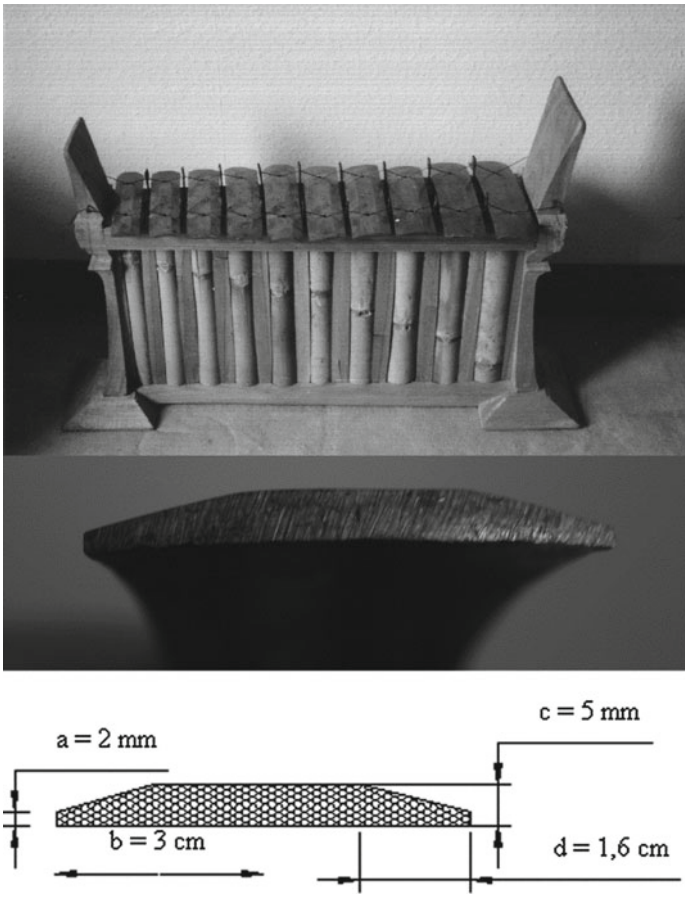


Fig. 1 Balinese *gender* (top), lowest plate front view of trapezoid shape (middle), and schematic view of trapezoid shape (bottom)

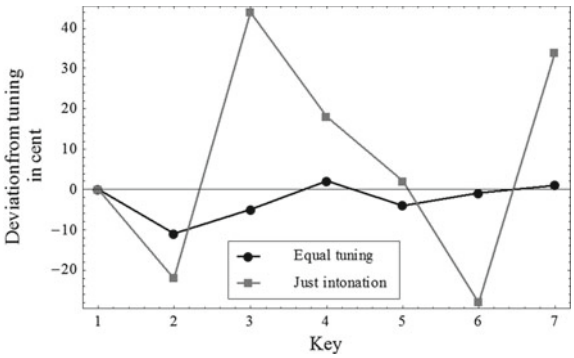


Fig. 2 Deviations of the tuning of a Cambodian *thake* from 7-tone equal tuning (black curve) and just intonation (gray curve). Clearly the heptatonic equal tuning is present in this instrument which cannot be retuned as the frets are fixed on the instrument body

This is seen when measuring the fundamental frequencies of the strings when playing the instrument. Table 3 gives the frequencies and cent for the *thake* investigated compared to the equal and just intonation, while Fig. 2 displays the deviations of the *thake* tuning from the two tuning systems. Clearly the equal tuning is much closer to the real tuning as is the just intonation. This is also not getting better when using for example a minor third with 316 cents instead of the major third with 386 cents. The 'third' of the *thake* with 342 cents is more a neural third and therefore does not also fit the minor tuning. Evidently the *thake* is tuned in a heptatonic equal tuning system known from Thailand.

8 Cambodian *Roneat Deik*

The situation is getting more complex when examining the *roneat deik* shown in Fig. 3. The one discussed here was collected at the Russian market in Phnom Penh

Table 3 Frequencies for one octave of a Cambodian *thake*

Fret	Frequency/Hz	Cent	Equal tuning/cent	Deviation equal tuning	Just/cent	Deviation Just inton.
1	173	0	0	0	0	0
2	190	160	171	–11	204	–22
3	210	338	342	–5	386	44
4	233	516	514	2	498	18
5	257	682	685	–4	702	2
6	284	856	857	–1	884	–28
7	314	1030	1028	1	1088	34

but built by an instrument maker near the capital. The builder explained to the author that it is common today for Cambodian musicians to retune their instruments either to the Cambodian or to the Western tuning. This became necessary as musicians are more and more often asked to perform with Western musicians and consequently the Cambodian musicians retune their instruments to fit the Western tuning. Tuning is done by adding wax to the metal bars, a common method throughout Southeast Asian. The Cambodian musicians are very well aware of the differences in the tunings and find it necessary to adjust their instruments to the respective scales. It is therefore interesting to examine the tuning of the present instrument which may give more insight into this tuning switch.

Table 4 gives the fundamental frequencies and cent relative to the lowest key for all 21 keys of the *roneat deik*. The octaves with respect to the lowest key are key 8 with 1209 cents and key 15 with 2407 cents, and therefore meet the perfect relation of 2:1 and 4:1 with 1200 and 2400 cents respectively quite good. Accepting these keys as octaves, we have three registers with seven keys in each of them. Figure 4 plots the cent values for all three octaves, taking the respective keys 1, 8 and 15 as fundamentals of these octaves. Clearly the pitches deviate slightly between the registers shown in Fig. 5. The differences between the octaves of the related keys show that register 2 (R2) and register 3 (R3) have similar deviations with respect to register 1 (R1) because the difference between R2 and R3 is generally smaller than that between R2–R1 and R3–R1.

Fig. 3 Cambodian *roneat deik* metallophone as used in this study

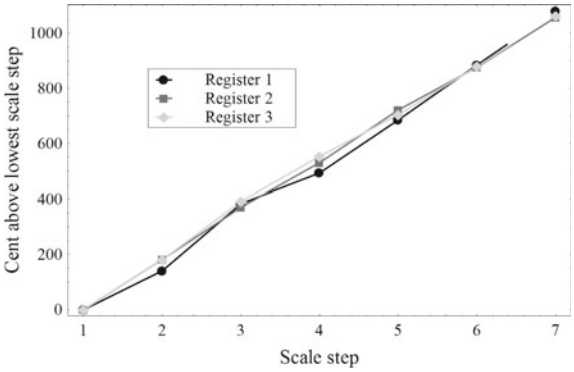


Fig. 4 Cent of scale steps between the three registers of a Cambodian *roneat deik* metalophone compared to lowest scale step of respective register

Table 4 Frequencies and cent values of the 21 *roneat deik* keys compared to 7-tone equal tuning, just intonation and a middle tuning between equal and just intonation

Key	Frequency/Hz	Cent	Equal tuning/cent	Deviation equal tuning	Just intonation/cent	Deviation just intonation	Middle tuning	Deviation middle tuning
1	176	0	0	0	0	0	0	0
2	191	141	171	−30	204	−63	187	−46
3	220	384	342	42	386	−2	364	20
4	234	494	514	−20	498	−4	506	−12
5	262	686	685	1	702	−16	693	−7
6	294	885	857	28	884	1	870	15
7	329	1081	1028	53	1088	−7	1058	23
8	354	1209	1200	9	1200	9	1200	9
9	394	1391	1371	20	1413	−22	1392	−1
10	439	1580	1542	38	1595	−15	1568	12
11	482	1740	1714	26	1707	33	1710	30
12	538	1931	1885	46	1911	20	1898	33
13	588	2086	2057	29	2093	−7	2075	11
14	653	2267	2228	39	2297	−30	2262	5
15	708	2407	2400	7	2409	−2	2404	3
16	786	2588	2571	17	2611	−23	2591	-3
17	889	2799	2742	57	2793	6	2767	32
18	976	2961	2914	47	2905	56	2909	52
19	1064	3112	3085	27	3109	3	3097	15
20	1177	3287	3257	30	3291	−4	3274	13
21	1309	3470	3428	42	3495	−25	3461	9

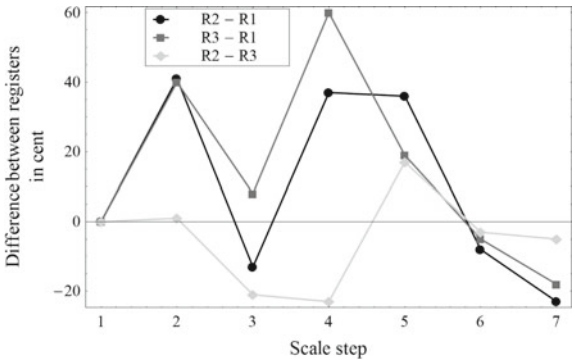


Fig. 5 Comparison between scale steps of three registers of a Cambodian *roneat deik*, register 2—register 1 (black line), register 3—register 1 (gray line), register 2—register 3 (light gray line). Register 2 and 3 are much more similar compared to register 1

As we are basically orientated about the registers, we can now compare the keys with the heptatonic equal tuning and the just intonation as shown in Fig. 6. The result is quite chaotic. Although there is a clear tendency for higher keys to show larger positive deviations compared to lower keys, no obvious tuning system can be observed. This may simply mean that the instrument is not well tuned and therefore useless for investigation. Still, as discussed above, the instrument builders report that they do not built the instruments according to one tuning but have in mind that the musicians will tune them in both directions, once to the heptatonic equal tuning and once to the just intonation. It seems worth it to compare the keys to a 'middle tuning' between both the equal and just intonation which is shown in Fig. 7. Interestingly a very clear picture appears when doing so. There are still deviations from this middle scale, but these are very systematic. The basic tendency for higher keys to deviate more than lower keys remains. Within the octaves, when again taking the octave keys 8 as basis of R2 and 15 as basis of R3, the pitches sharpen within one octave at first, up to the fourth key, only to then flatten to the next octave pitch. Only the second key in R2 and R3 flattens at first, still consistent between R2 and R3. This pattern continues for the higher keys in R1, not over the whole R1 but only within keys 4–7. In conclusion, we may argue that R1 is not perfectly in line with this reasoning.

Interestingly the tuning pattern shown in Fig. 7 is systematic for R2 and R3 and corresponds to the idea of building an instrument which may be tuned in two directions, to the heptatonic equal scale and to the just intonation. This can also easily perceived when comparing the *roneat deik* with the *thake* and with just intonation (Comparison_ThakeRoneatDeikJustIntonation.wav). In the demo first one octave is played on the *thake*, the *roneat deik* starting at plate 8 and a synthesizer sound toned to just intonation. Then the three are compared for each key. The *roneat deik* as the middle sound is most often clearly higher in pitch than the *thake* but lower than the just intonation.

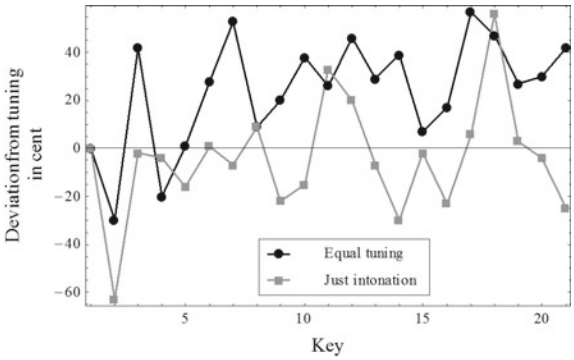


Fig. 6 Deviation of key tunings of Cambodian *roneat deik* from heptatonic equal tuning (black curve) and just intonation (gray curve). No clear pattern appears in both cases

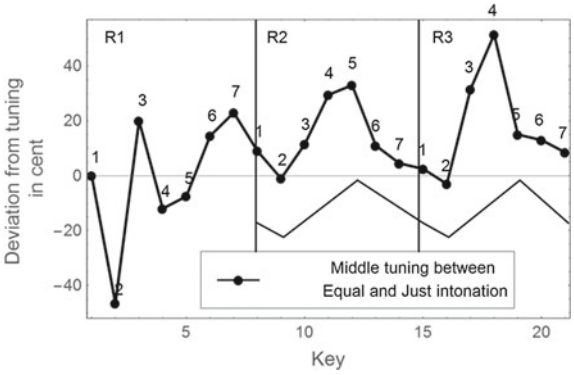


Fig. 7 Deviation of key tunings of Cambodian *roneat deik* from a tuning which is in the middle between heptatonic equal tuning and just intonation. Here a regular pattern appears with regular octaves, and a systematic tendency to sharpen the pitches both within an octave and over the three registers

The sharpening of higher keys is very well known from Western tunings, especially from the piano which also sharpens the higher keys to add a bit of tension in the tuning, making the sound more interesting.

9 Cambodian *Smot*

The Cambodian Buddhist chanting style *smot* is special as it is very melismatic [26]. Monks train for years and competitions are held. When investigating tunings, singing seems to be a natural choice as the voice can be freely intonated and therefore should display the tuning system used. Figure 8 shows one phrase of a *smot* tune *Bat*

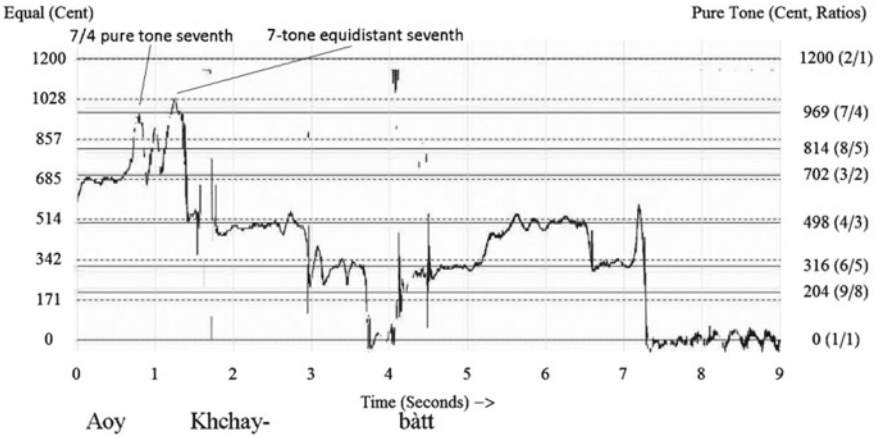


Fig. 8 Phrase 14 of Cambodian *smot* chant *Bat Sara Phanh* as sung by Kai Sokmean (see sound file 5_SARA PANH_Phrase_14.wav for only this phrase). The overall pitch variations caused by melismatic chanting is generally larger than the difference between just intonation (thick horizontal lines, cent on the right) and equidistant scale (dotted horizontal lines, cent on the left). The deviations between the seventh which is 7/4 and 969 cents in just intonation and 1028 cents in equidistant scale are performed at the beginning of the phrase with peaks meeting the two tunings precisely, causing a mixed impression of the two tonal systems. From Bader 2011

Sara Phanh as performed by Kai Sokmean at the *Moni Prosity Vong* monatry near Phnom Penh recorded by the author in 2010 (see sound file 5_SARA PANH.wav). Kai Sokmean won the *smot* contest of the area the year before the recording. As shown before with the *thake* and the *roneat deik*, the heptatone equidistant scale is performed in Cambodia, but the Western scale is very well known and musicians are aware of both by retuning their instruments to one of the two according to the music performed. The melismas of *smot* are trained for many years where the placements of vibrato and pitch glides are fixed. Thus we can expect the intonation to be precise to a certain amount. Performance intonation by a highly skilled singer as investigated here is expected to be intended.

Figure 8 shows phrase 14 of *Bat Sara Phanh* in comparison to two heptatonic scales, a just intonation (thick horizontal lines, right scale) and an equidistant scale (dotted horizontal lines, left scale). For the most part of the phrase, the vibrato width is larger than the difference between both tuning systems and the trend continues with all other phrases investigated. In the beginning of the phrase, a melisma is performed, lasting for about a second, where the first and the third peak of the intonation meet just intonation and the equidistant tuned seventh precisely. Indeed the largest deviation in intonation between the two tuning systems appears with this seventh and an intonation difference will easily be perceived. Although it is a subjective judgment, the author perceives this phase as a mixture of both tuning systems and with the pleasure of witnessing a highly virtuous performance.

10 Sun Pyi (*Bamboo Flute*) and Dum Ba (*Zurna-Type*) of *Kachin* Wunpawng *Band*

The Kachin State of modern Myanmar officially consists of six tribes: Lishu, Jingpaw, Rawang, Maru, Lashi and Azi [27]. The name Kachin is used by the Myanmar officials as a collective name but was introduced even before colonial times by the Bama, the largest ethnic group in Myanmar where the British derived the name Burma from, and by the Shan, an ethnic group living in the mountains along the Thai border [28]. The civil war between the Kachin and the central government since 1961 formed a Kachin Independence Army (KIA) and so the different tribes identify with the Kachin name, although people still have two names: a Kachin name and one of their ethnic group. The identity as Kachin also appears in the *manao*, a New Year celebration with a central feast in Myitkyina in low-land Kachin state. Furthermore, the Kachin present themselves in the Kachin National museum also located in Myitkyina. There, the different tribes with their costumes, habits and musical instruments are displayed as well as the Frazer alphabet, which was invented for the Lisu in 1918 by J.O. Frazer, a Scottish missionary. In what is probably the first printed language handbook by Hertz [29], the Kachin are identified with the Jingpaw (Chingpaw) pointing to the largest group within the Kachin [29]. Still the Kachin derive themselves from the sixth son (*pawng yaw*) or tribe of the Mongols and call themselves *wunpawng*. Therefore they are different from the Bama or other ethnic groups from the Tibeto-Burmese lineage or the Mon/Khmer groups. The ethnic groups sometimes also associate themselves with larger cities. It can be seen from the traditional musicians head dressing, a bound cap, which is yellow for musicians from Bhamo, a town south of Myitkyina, and red/blue for those from Myitkyina.

Basically the low-land Kachin around Myitkyina are Baptists and the mountain Kachin around Potao, near the Chinese border, are more animistic. The low-land Kachin also have many animistic habits like fortune telling from smoke [30]. In the mountain region, drums [31], slit-drums [32], gongs [33], flutes, mouth organs and stringed instruments [34] are in use. Songs and dances are performed for several occasions such as before crop harvesting or when beginning a long journey. The author recorded a performance of Myittung Gam (Lisu name: Dawng Da) playing the *hkibu* (ƆI Bך in Frazer alphabet), a three-stringed fretted instrument (open string tones C–F–C), while dancing and singing after planting a new tree.

The low-land Kachin music is not very prominent in everyday life. For feasts such as Sunday Baptist service, house-warming feasts, weddings, funerals or for the *manao* New Year feast, *wunpawng* drum-and-shawm-like marching bands perform, consisting of *sun pyi* (flute), *dum ba* (zurna-type), *bou* (gong), *shup sheng* (cymbals) and drums (Western Bass Drum and tom-toms). Other musical performances are church choirs performing with reed organs. In addition, more and more young people play Western guitars which are built in Myanmar and perform the Bama based musical style called stereo [35], where mostly Western pop and country and Western style songs are performed in a Western band setup still with texts in local languages. Videos showing performances of *manao* with the original music replaced by playback or

Fig. 9 *Masum wunpawng* (*wunpawng* trinity band) performing at Kaiwa Kha village, Kachin state, Myanmar

karaoke music are mostly soft synthesizer music with local melodies and singing. Regional artists perform either more traditional music or more Western oriented styles, nearly always with a soft synthesizer background and soft drum machine sounds. The commercially available music is thus considerably different from the traditionally performed music of *wunpawng* bands, mouth organ playing or folk songs.

10.1 Wunpawng Dum Ba *Shawm Tuning*

The *dum ba* shawm played in the *wunpawng* band in Fig. 9 by the very right musician has four finger holes at the front and one for the thumb at the back. In the recordings, up to nine tones were played with many melismatic ornamentation like trills and slides. Figure 10 shows the deviations of the notes played from the nearest pitches of the equidistant and of the just intonation. The fundamental is taken where the melody has its resting point. With the just intonation, a pattern appears where the fundamental, the lower major sixth and the upper fifth are in good alignment so that they come close to a just intonation relative to one another. This also holds for the lower fifth and lower seventh, upper fourth, upper seventh and the second above the octave. Still both rows are set about 40 cents apart. The octave is stretched and does not belong to one of these series. Generally, a tendency to decrease pitch with higher notes is also obvious. The deviation pattern for the equidistant scale does not show such a clear pattern.

It was discussed above that the tuning of wind instruments is a demanding challenge for instrument builders since the tuning depends on the bore, the horn, the tube end-correction as well as the finger hole depth and size [10, 11]. As there is no way to tune a wind instrument perfectly, players need to adjust the pitch by adapting their blowing pressure and lip tension, sometimes also the partial covering of a finger hole. The pitch played is a compromise between the abilities of the instrument and the player's intentions while players need to accept mistunings to a certain amount. Therefore instrument builders often place the finger holes according to the pitch of a reference tone. Then the tuning system does not necessarily have all pitches of a theoretical scale but fit to the pitch it is referenced to when building the instrument.

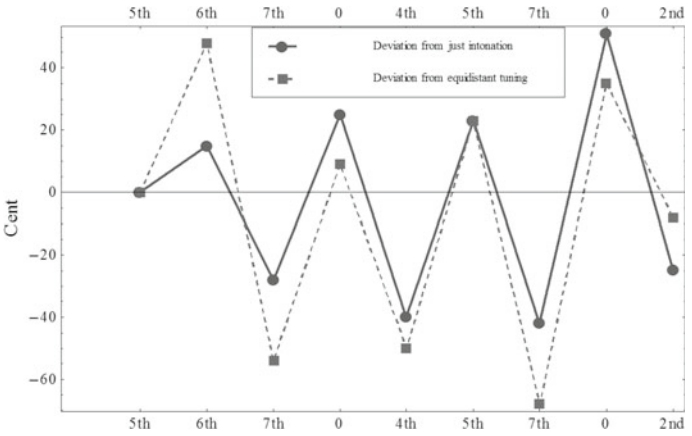


Fig. 10 Deviations of *wunpwang dum ba* double reed instrument from just intonation and equidistant tuning. The deviations to just intonation show a regular pattern

The author asked the *dum ba* player to play a scale from the lowest to the highest note. Although the musician was trying to play such a scale several times, he did not succeed. He obviously never had practiced a scale but only the melodies. From this and the findings of Fig. 10, it can be reasoned that the tuning is built of relative intervals, here fourth, fifth, and third intervals, rather from a global tuning approach. Although the player adjusts the pitches while playing, he still needs to deal with the pitches present in the instrument making compromises or temperaments. Of course, the melodies played had both intervals within one row as well as intervals between rows as all pitches were present in nearly all tunes. So it seems that the player tried to temper the pitches available on the instrument to arrive at just intervals if possible.

Such a tuning strategy is similar to the idea behind a Renaissance mean-tone tuning where several just third intervals are intonated with the trade-off that other intervals are widely out of tune. Such a tuning does not take the global view of tempering all intervals but favor a few and accept rough detunings with others. This tuning strategy has also been suggested with European bagpipes where just fifth are present while other intervals are very much out of tune in terms of just tuning. This is considerably different from the Baroque tuning systems such as Werckmeister, Kirnberger, Valotti or Young, which take a global view of making all intervals playable. In the tuning system of the Kachin, there seem to be some intervals of higher importance than others, which therefore are tuned fairly well in terms of a just tuning by accepting that other intervals are to be widely detuned. In the *dum ba* tuning, the octave is the only exception which again points to the importance of this interval.

Fig. 11 Wind instruments with equidistant tone hole spacing. From top to bottom: *hne* (Sri Lanka), *suling* (Bali), and two *sum pyi* (Kachin)

10.2 Wunpawng Sum Pyi *Bamboo Flute Tuning*

The tuning of the *sum pyi* bamboo flutes seem to be more complex when looking at their design in Fig. 11, where the flutes are the two lowest ones, collected by the author from the *wunpawng* band musicians in Myitkyina in 2013 (Kachin_WunPawng_Band_3.avi). The tone hole distances are equal and hence the tuning is expected to be neither equidistant or just. Such flute fingerhole spacing can be found all over Southeast Asia; other examples are a Balinese *suling* flute collected by the author in Bali in 1999 and a *hne* zurna-like instrument collected by the author in Sri Lanka in 2014 and discussed below. The author witnessed the building process of the *suling* where the builder took a bamboo piece as a measure between the tone holes, therefore intentionally making all tone hole distances equal.

The *sum pyi* flute players played different tunes from which the pitches were analyzed. Additionally, the author asked the flute players to play a scale on these instruments and the musicians were able to do so again with some trouble, especially when playing the fifth note in an upward direction (see video Kachin_WunPawng_Band_Flute_Scale_2.avi). The *wunpawng* musicians know the Western solmization and use it when referring to notes in the performance. The tunes played were pentatonic while the scales played had seven steps in the octave. Also the tunes had the final or main note at the second step of the scale played which is used as a reference pitch for the tunes.

As shown in Table 5, the equidistant tuning is not reached at all with the flutes. The deviations of the performed scale go up to a semitone (104 cents) in step 3 and step 2 is 54 cents apart. The octave has a deviation of 10 cents, while the nearest pitch is step 5 with -3 cents. So the *sum pyi* scale is too far off to be called heptatonic equidistant. This also shows up in the performance. Interestingly the deviations are very different between scale and performance in both cases of just and equidistant

Table 5 Deviations in cent from nearest pitches of equidistant and just tuning and pitches played by *sum pyi* player of a *wunpawng* band once during a song performance (Performance) and once when playing a scale (Scale). As the song is pentatonic, two steps are missing which are step 2 and step 5 in equidistant tuning, in terms of just intonation the third and sixth in just tuning. The main note in the performance is the second lowest pitch, therefore this has been taken as a reference

Scale equidistant	Step 0 0	Step 1 22	Step 2 56	Step 3 104	Step 4 17	Step 5 −3	Step 6 10	Step 7 −12
Scale just intonation	Prime 0	Maj 2nd −11	Maj 3rd 12	Tritone 28	Fifth 0	Min 6th 38	Min 7th 20	Octave −12
Performance equidistant	Step –1 –32	Step 0 0	Step 1 73	Step 325	Step 4 −12	Step 6 -47	Step 7 −23	Step 8 31
Performance just intonation	Min 7th –22	Prime 0	Maj 2nd 40	Fourth 41	Fifth −29	Min 7th −37	Octave −23	Maj 2nd −2

tuning again pointing to the wide variability a flute player has to tune pitches using articulations.

Comparing the scales to a just tuning, the deviations are considerably smaller; especially the fifth which is met perfectly. This note was produced with much care by the musician, that is with much effort in terms of adjusting the playing pressure and the distance between mouth and mouthpiece of the flute, two parameters a player can use to adjust the pitch of a flute as discussed above. Still the scale coming closest to the pitches played has a major second and third but also a minor sixth and seventh. Clearly the equidistant tone hole spacing, which suggests an equidistant scale, leads to a sixth and seventh closer to the minor steps of a just tuning rather than to the major steps. Instead of a fourth, the pitch played comes closer to a triton which is also closer to the equidistant finger hole placement of the instrument. Therefore it seems that the player tried to meet just these pitches. Obviously the player tried to meet a just scale with his instrument when asked to play a scale.

This adjustment to just tuning is not clear anymore in the performance. Although the deviations for a just tuning are still much smaller than that for an equidistant tuning, there no longer is a perfect fifth, and the major second and the fourth are deviating quite strongly. The pitch played is closer to a fourth than to a triton and so the just tuning appears stronger here. Clearly the players strongly change the instruments playing parameters such as the blowing pressure or the distance between the mouth to the mouthpiece to adjust the tuning to a desired one. They obviously try to meet a just tuning and try to match the pitches as good as possible when playing a scale. When performing a piece, the just intonation comes closer to the pitches played than does the equidistant tuning.

The question arises as to why the instrument is built with equally spaced finger holes. Although there is no clear conclusion to draw from the present data, the reason could be the simplicity of the building process. All flutes were built by the musicians themselves and placing finger holes equidistantly is the simplest arrangement possible. As the player needs to adjust the pitch later anyway, this equidistant arrangement of finger holes is a good compromise, a reasonable temperament. Flutes with such finger hole distances are often found in Southeast Asia. The building process of the

Balinese *suling* flute displayed in Fig. 11 was witnessed by the author. The builder, a peasant near Ubud, took a short bamboo stick and cut it to the appropriate length of the desired distance between two tone holes. After cutting one hole in another bamboo tube, he placed this stick at the end of this hole and the other end of the stick then determined the position of the next hole. After repeating this process, all tone hole distances were equal. The building process of the whole flute took ten to 15 min and such flutes can easily be replaced when broken. It is therefore reasonable that for such a simple instrument, no complex tuning process takes place and hence it is up to the player to tune the instrument on the fly while playing.

11 Burmese *Pat Wain* Drum Circle Tuning

The *pat wain* is a drum circle consisting of about 22 pitched drums so that melodies can be played. It is used in *hsain wain* music in Myanmar for *zat pwe* entertainment and puppet shows or plays [36]. A pitched drum set of this size does not exist elsewhere, yet pitched drums are known to some extend in Western Rock and Jazz tom-toms. Enlarging the amount of drums makes melodic play possible, maybe most prominently performed by the American Jazz rock drummer Dave Weckl. The Cuban *conga* drum pair also has two different pitches which probably was the invention of Carlos Valdes, a Cuban musician who invented the hoop in order to change precisely the tension of the drum head using tension rods [37]. The Creole *conga* from Africa was low pitched and only during the invention of Cuban *son* music, starting from *changüi* around Guantanamo, the drums were pitched high to interact with the *tres* and the vocals. Then the drums were tuned by heating, which is not precise enough to tune accurately. Only after the invention of the drum hoop, two carefully tuned pitches could be produced. The *pat wain* is tuned using a paste of rice and ash as usual with drums in Southeast Asia. By applying the paste on the top of the drum head, the pitch can be slightly increased or decreased and therefore the pitch can be tuned very accurately. Unfortunately, this tuning process takes some time and *pat wain* players, who are normally the head of the *hsain wain* ensemble, have sometimes been entertaining the audience with short stories or jokes while tuning from one Burmese tuning system to another, as an example is known from a famous *pat wain* player Sein Beda [38].

As discussed above for the Kachin *dum ba*, tuning processes can clearly decide the resulting tuning. Tuning processes for another instrument found in Myanmar, the Burmese harp *saun gauk*, which is mainly tuned according to octaves and fifth is reported by Williamson [3]. Bama music theory knows three main tone steps which range from top to bottom like *tya, tei, tyaw*; C, B*, and G in Western notation. The star indicates a neural interval between a major and minor seventh. The steps for tuning the *saun gauk* which Williamson reports are

(1) Start with C4 (*tya*) and tune the lower octave C3, then tune B*3 (1050 cents) freely, from B*3 tune E*3 as a fifth, then from C3 tune G3 a fifth higher to end with the three pitches t*ya, tei, tyaw* (C4, B*3, G3).
(2) From C3 tune F3 as lower fifth.
(3) Tune additional octaves E*3 ->E*4, F3 ->F4, G3 ->G4, B*3 ->B*4 and the other octaves.
(4) Finally, from G4 tune D5.

Nearly the whole tuning process is done via octaves and fifth, the intervals most easily tuned by ear.

The instrument investigated is played by Kyaw Zay, who is leading the group *Kyaw Zay hsain wain*, and was recorded by the author in Myitkyina in 2013 (see video Mytinya_PatWain_Performance.mp4). Kyaw Zay tuned the instrument extensively before performing (see video Mytinya_PatWain_Tuning.avi). As reported by Williamson on the *saun gauk*, the tuning is mostly done using octaves, fifth and third intervals to tune the drums. This tuning lasted for half an hour before the player was satisfied. The drums were recorded individually leaving enough time for the whole sound decay. This is necessary as drums may have pitch glides where the pitch at the beginning of the sound is higher than at its end, especially with the low-pitched drums. This pitch glide might last 100–300 ms. Generally the overtone structure of a single drum sound is a mixture between a harmonic and an inharmonic overtone. The general solution of the radial wave equation which holds for round membranes calculates some eigenvalues which are very close to a harmonic series of 1:2:3:4 [39]. The other partials of a drum spectrum are inharmonic to this harmonic series. Therefore drums can be used as a pitched instrument to a certain extend.

Figure 12 shows the deviations of the investigated *pat wain* from a just tone and from an equidistant scale. Clearly the just intonation fits better with fewer deviations. Again the musicians know and use the Western solmization. This is the case in nearly all Southeast Asia. For Myanmar, the reason could be the introduction of Western notation in Burma at the beginning of the twentieth century since traditional music in Burma did not have any notation. The low Do was used as a reference, the octaves Do′ and Do″ do show only slight deviations and a slight octave stretching, known also from Balinese music [7] or Western piano tuning [16] and in many other cases around the world. When examining the pitches between the octaves Do-Do′ and Do′-Do″, a similar pattern appears, namely a flattening of pitch with the next note Mi and Mi' followed by a sharpening of pitches from Fa-Ti and Fa′-Ti′. This pattern was already found with the *roneat deik* discussed above in nearly the exact same way. There might be a preference for such a tuning in terms of making the scale more interesting with these small deviations, which is a similar reason as for octave stretching. There also might be another constraint not yet known to the author with the need of tempering the tuning accordingly.

The pitches Do″-Ti″ are not following this direction. These drums decay very fast and it is hard to determine a clear pitch here anyway. Also the additional inharmonic overtones appear stronger with higher pitched drums and a pitch sensation is weakened again. The low pitches SO-Ti need to be treated separately, too, as they have a

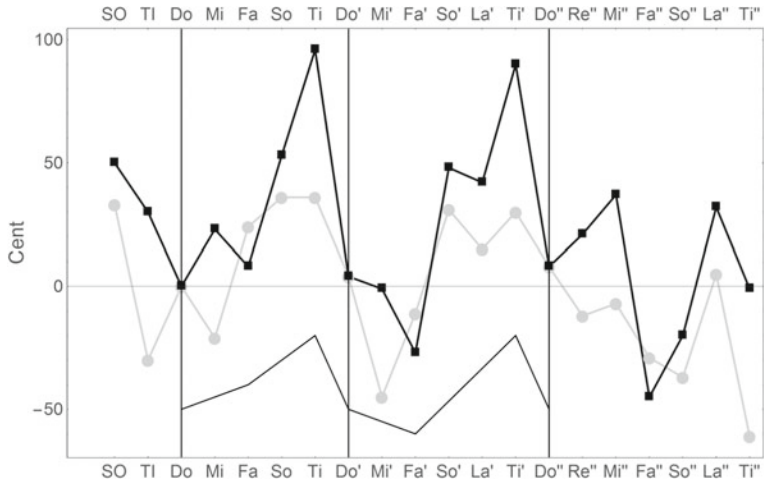


Fig. 12 Deviations of the twenty drums of the Kyaw Zay *pat wain* from just intonation (circles) and equidistant tuning (squares). The just tuning fits better, where the octaves are very accurate, and the scale between the octaves shows a pattern of a flattening with the lower notes and a sharpening of the higher ones within one octave, a pattern very similar to that of the Cambodian *roneat deik*

considerable pitch glide at the tone onset lasting from around 320 ms with SO and 150 ms with Ti. Figure 13 shows the differences between the pitches at the beginning of the tone and that of the decaying drum. The pitch glide can go up to nearly 500 cents which is a musical fourth! The second integration time of the ear, within which perceptually all auditory data are present as one event and no further temporal division can be done, is 50 ms. Therefore the pitch of these drums decays are clearly audible and with fast playing, these pitches are often quite arbitrary. Both restrictions, the fast decay and lost pitch perception with high pitched drums and the pitch glide with low drums in connection with the mixture of harmonic and inharmonic overtone structures, might be the reason why drum circles are not found very often as melody instruments. In terms of the tuning system discussed here, it is important to note that the pattern found in the two octaves between Do and Do″ do not need to be repeated in lower or higher octaves as pitch perception is not so clear with a drum circle.

Burmese music knows several of the tuning systems discussed above, so it is interesting to calculate the standard deviation of the measured pitches of the *pat wain* with the different tunings shown in Table 6. The lowest deviations are for the just intonation as well as for the traditional *hyin lon* tuning. This means that the deviations from just tone discussed above are in such a way to make the tuning as close to just tone as to *hyin lon*. Once again, it appears that a tuning system tries to balance between just tone and traditional tuning in a way to produce the lowest deviations to both systems.

As the timbre of the *pat wain* is a mixture of harmonic and inharmonic overtones, the Helmholtz idea of a tuning system based on intervals showing least roughness

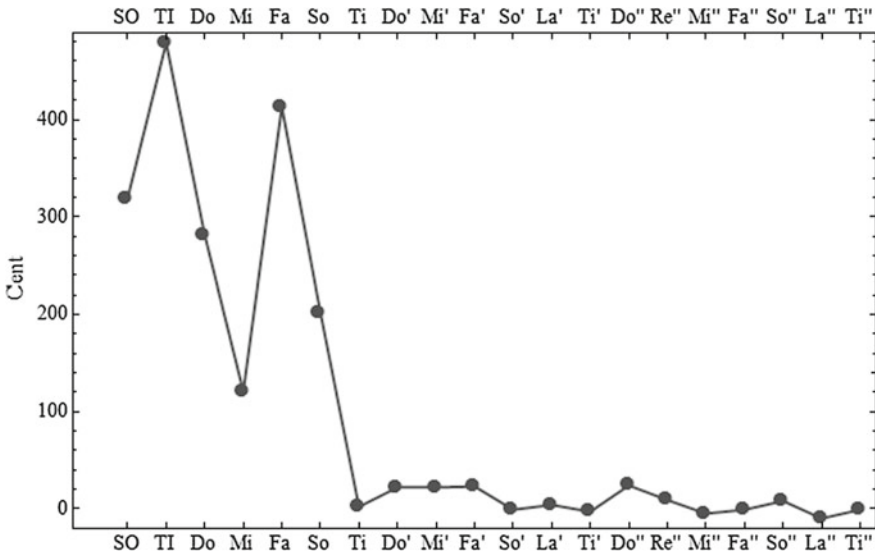


Fig. 13 Pitch shift in cent for *pat wain* drums between the beginning of the tone and its steady sound after 320 ms with SO and 120 ms with Ti. The deviations are strong up to So and only slight thereafter

was tested with the *pat wain* in two ways. As the instrument is played by two hands often in a highly virtuosic way, two possible pitch combinations were tested. First, simultaneously played octaves were mixed from the recorded single-tone sounds between Do-Do′, Mi-Mi′, Fa-Fa′, So-So′, and Ti-Ti′. Secondly, again from the original recorded tones, two simultaneous tone pairs Do′-Mi′, Do′-Fa′, Do′-So′, Do′-La′, and Do′-Ti′ were built. These pairs of tones are at the original pitches as recorded in the field. Then both tone series were produced again but this time with the higher pitches adjusted. In the octave series, the higher pitches were adjusted to fit a perfect 1200 cents octave. In the second series, the higher pitches were again adjusted to meet the just intervals of 368, 498, 502, 886, and 1018 cents for major third, fourth, fifth, major sixth and minor seventh which came closest to the recorded pitches. The adjustments were performed by time stretching to maintain the sound itself. As the

Table 6 Standard deviations between the *pat wain* tuning of Kyaw Zay and the different tunings. The lowest values are found for just intonation and for *hyin lon*, the traditional Burmese tuning. As both values are the same, the tuning tempers between just intonation *hyin lon* tuning and therefore is a real temperament

	Hyin Lon	Auk Pyan	Pale	Myin Zain	Just intonation	Equidistant
Standard deviation /cent	29	38	47	49	29	35

Table 7 Roughness calculated by the Sethares and Helmholtz/Bader algorithms for the *pat wain* drums tone pairs with original and just intonation adjusted pitches in series of octaves and intervals. As the algorithms are differently scaled, the absolute values need to be different between the algorithms. Clearly the differences of roughness between the original and adjusted pitches are small and all adjusted pitches show less roughness compared to the original ones. So the idea of non just tunings of instruments with inharmonic spectra caused by least roughness of their partials fails with the *pat wain*

	Do-Do′	Mi-Mi′	Fa-Fa′	So-So′	Ti-Ti′
Played Sethares	0.50	0.80	0.45	0.59	0.06
Adjusted Sethares	0.46	0.34	0.31	1.22	0.08
Played Helmholtz/Bader	3.15	4.6	2.51	3.29	0.28
Adjusted Helmholtz/Bader	2.93	2.00	1.70	7.01	0.37
	Do′-Mi′	Do′-Fa′	Do′-So′	Do′-La′	Do′-Ti′
Played Sethares	0.14	0.08	0.072	0.009	0.009
Adjusted Sethares	0.12	0.076	0.065	0.009	0.006
Played Helmholtz/Bader	1.06	0.74	0.58	0.09	0.82
Adjusted Helmholtz/Bader	1.00	0.71	0.56	0.09	0.03

shifts were only slight, with a minimum shift of 5 cents to a maximum of 54 cents, the sounds still sounded very realistic.

For both series, the roughness of the sounds was calculated, once with the algorithm of Sethares and once with a reimplementation of the original Helmholtz algorithm (Helmholtz/Bader). As Helmholtz did not give a mathematical formula, the algorithm was built to have a maximum roughness at a frequency difference of 33 Hz and to decay with higher differences. The algorithm was tested with tuning systems and fit well with expected values [40]. The results of both algorithms differ in absolute values but behave very similarly in terms of the order of more or less roughness of the sounds. Table 7 shows the values for both cases. With the second series, roughness clearly decreases with higher intervals as expected. Still roughness itself is very low in absolute values. The differences between the original and the adjusted scales are very small and barely perceivable. With the octaves, roughness is larger mainly because lower-pitched *pat* (drums) were involved. Except for one case of higher roughness, the difference between both cases is small but may be audible. In both cases, the roughness of the adjusted cases is always higher than that of the original. So the reasoning that the deviation of the tuning from just tuning is used to minimize the roughness of inharmonic overtone spectra clearly fails with the *pat wain*.

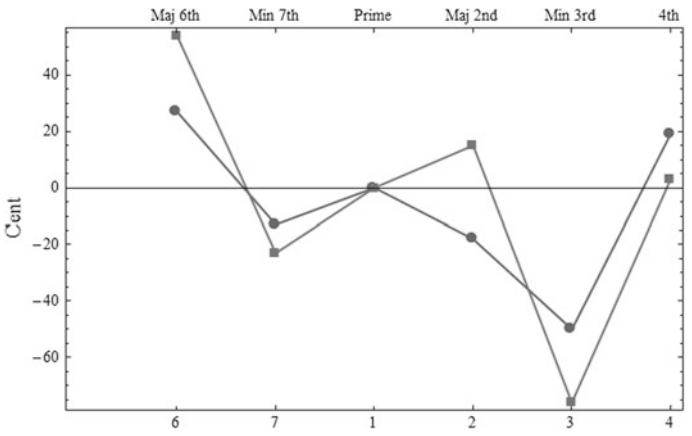


Fig. 14 Deviations in cent from just intonation (circles) and equidistant scale (squares) of six pitches of a Singhalese *hne* (zurna-type instrument) played during a *puja* at the Tooth temple in Kandy, Sri Lanka, recorded by the author in 2014. Although the *hne* has an equal spacing of finger holes, the pitches deviate strongly from an equidistant scale and come closer to a just tuning

12 Singhalese *Hne* Shawm Tuning

Another example of a wind instrument with equally spaced finger holes is the *hne* played in south Sri Lanka at Singhalese Buddhist *puja* rituals (see video Kandy-ToothTemple_hne.avi). The present example is a recording made by the author in 2014 at the Tooth Temple in Kandy during the daily ceremony. The performance is again a typical drum-and-shawm ensemble where the *hne* is performing as the only melody instrument. Figure 14 shows the deviations in cent between the pitches played during the performance and the just tone (circles) and equidistant (squares) tunings. The base tone of the melody is denoted as prime or step 1 and for the just scale case, the just scale steps closest to the measured pitches were taken. It clearly appears that the tuning strongly deviates from an equidistant tuning. Furthermore, the referenced pitches of a just scale come even closer to the pitches played. The deviations for the second step are about the same for both tunings and the fourth step is a bit closer with equidistant tuning. The third, sixth and seventh steps are considerably larger for equidistant than for just tuning. As discussed above, the bore of the *hne* tube, the horn at the tube end as well as the blowing pressure change the instrument pitch considerably. So, again, equally spaced finger holes on the instrument do not mean that the pitches played come close to an equidistant or a just tuning.

13 Summary and Conclusions

The examples discussed above support in many ways the idea of Southeast Asian tuning systems caused by compromises between constraints, therefore being real temperaments.

- Cambodian *roneat deik* builders report the necessity to retune their instruments either to their native scale or to a Western scale when playing with Western ensembles. Measurements of the *roneat deik* find this compromise implemented in its heptatonic scale.
- The *pat wain* drum circle tuning comes closer to a just tone than to an equidistant scale, still without meeting the just tone perfectly, which would easily be possible. The detuning from just tones follows a pattern also found with the Cambodian *roneat deik* and therefore seems not to be random but intended. Also the detuning from just tone is not caused by a decrease in roughness, which would follow the Helmholtz idea, since retuning to a just scale decreases roughness.
- The *pat wain* tuning is a compromise between just tone and traditional *hyin lon* tunings.
- The heptatonic scale of the Kachin *wunpawng* band as played on the *sun pyi* flute built with equidistant spacing of finger holes deviates largely from an equidistant tuning and is closer to the just tuning.
- When asking *sun pyi* musicians to play a scale, they very much try to meet just pitches by adjusting the instruments playing parameters.
- The Kachin *dum ba* shawm instrument is obviously tuned in two rows, where each row closely follows a just tuning on its own. The difference between the two rows seems to be caused by the problem of tuning wind instruments and the tuning procedure of tuning octaves and fifths.
- The Singhalese *hne* shawm instrument has equal spacing of finger holes and again is played closer to a just tuning than to an equidistant scale.
- Cambodian *smot* singing is very melismatic with much vibrato of width mostly larger than is the difference between just and equidistant tunings. Still in the melisma discussed above both seventh, the just and the seven-tone equidistant seventh are met perfectly.

As the musical discourse in Southeast Asia was strong over the centuries, such compromises were needed. Williamson reports on the adjustment of the Bama scale to the Thai scale after the defeat of Ayutthaya to the Bama (1767), which was followed by a strong interest of Bama people in the Thai plays of Ramayana played in heptatonic equidistant scale which made adjustments from *hyin lon* scale necessary to *auk pyan* or *pale* scale. The influence of the Ottoman Janissary bands, as displayed by the many Janissary-like marching drum-and-shawm bands discussed above with the Kachin *wunpawng* and the Singhalese *hne* temple music, strongly suggests such adjustments of the tuning system. Trading routes might also have enhanced musical interchange. The special nature of *Saṅgītopaniṣat-sāroddhāraḥ* music theory with in-depth discussion of melismas and Tantric features of tones and melodies was

probably enhanced by the trading on the west coast of India where this treatise is assumed to originate from.

Such temperaments or compromises could also be the reason for equidistant scales, where the most prominent ones are

- five tones: Indonesian *sléndro* scale,
- seven tones: Thai scale also found in Cambodia and elsewhere in Southeast Asia,
- twelve tones: Western equal tuning, and
- 22 tones: ancient Indian scales (*Natyashastra, Cilappatikāram, Saṅgītopaniṣat-sāroddhāraḥ).*

The twelve-tone Western equal tuning is known to be such a compromise. Following the interpretation of Rowell of a 22-tone equidistant *śruti system with the Natyashastra* and *the Cilappatikāram*, also present in the *Saṅgītopaniṣat-sāroddhāraḥ*, as a flexible switching between modes, it appears that the ancient Indian music theory has applied the idea of a compromise between pitch played and just tuning much earlier than Western music theory as suggested from Renaissance mean-tone tuning and Baroque time by Werckmeister, Kirnberger, Valotti or Young. If the pentatonic and the heptatonic equidistant scales are also such a compromise cannot be answered in historical terms. As discussed above, musicians tend to obtain a just scale if possible. In light of the findings of musicians trying to temper tuning systems along constraints, necessities in the past might have led to the development of these scales.

From the measurements and interviews discussed above, one can have no doubt that the tuning systems of the instruments as played today are caused by compromises between constraints. It is no surprise that when the instruments are measured by themselves the tunings are often found not to follow an existing tuning system and therefore might be considered as arbitrary. Still if a non-Western musicologist would investigate Western instrument tuning, he would find stretched octaves with the piano, a special tuning of the guitar caused by differing material dependent string tension when pressing the strings down not leading to a twelve-tone equidistant scale, or very strange tuning systems with wind instruments here again caused by the complexity of wind instrument tuning. When only taking these measurements into consideration, the tuning system of the West would not be straightforward to find either. Many treatises discuss all kinds of systems and therefore it is possible to understand tuning systems by both measurements and theory. In Southeast Asia, there are very few written sources and musicians as well as instrument builders are often not aware of tuning systems but only perform in a traditional way – indeed a situation also not too far off from Western habits of musicians and instrument builders.

To summarize, the main reasons to deviate from a desired, most often just tuning are

- the necessity of a compromise when playing with the same instrument set in differently tuned ensembles,
- problems in instrument building, especially with wind instruments, or
- tuning deviations caused by inharmonic overtone spectra, short-lived tones (like with high pitches of the *pat wain*) or pitch glides intrinsically in the tones.

The compromise found today between traditional and just tuning systems cannot be understood in terms of the increased presence of Western music in Southeast Asia according to the discussed findings that there is not *a* Western tuning. So with the above findings and following the reasoning of universals and temperament as a system of constraints, the main constraint still seems to be a just system. Therefore it is much more likely that just is indeed a universal of music all around the world which is often intended but also often deviated from because of other constraints discussed above, such as instrument building or playing abilities.

Acknowledgements My thanks to the performers and informants in Cambodia, Myanmar and Sri Lanka, especially Kai Sokmean So Tia, Savuth Prum, Prof. Dr. Annang, Kyaw Zay, Na Na Masan New Songi, Hpung Taung La Hum Seuy Aung, Myittung Gam, the UNHCR people Paul Knudsen, Lu Zaw and Saw Yu for their wonderful music, support, ideas and patience. Also many thanks to Anton Isselhardt in Phnom Penh, whose excellent work and engagement in music, organization of music festivals, and musical exchange between Cambodia and Germany for many years enriches the musical scene in Cambodia a lot, and to my dear colleague Dieter Mack for his links and support. Also many thanks to all the others helping at and around the field works.

References

1. Ellis AJ (1884) Tonometrical observations on some existing nonharmonic musical scales. Proc Roy Soc 37:368–385
2. Schneider A (2001) Sound, pitch and scale: From “tone measurements” to sonological analysis in ethnomusicology. Ethnomusicology 45(39):489–519
3. Williamson M (2000) Burmese harp, vol 1. Nothern Illinois Monograph Series of Southeast Asia, Illinois
4. Garzoli J (2015) The myth of equidistance in Thai tuning. Anal Approaches Music 4(2):1–29
5. Sethares WA (1998) Tuning, timbre, spectrum, scale. Springer, London
6. Vetter R (1989) A Retrospect on a century of gamelan tone measurements. Ethnomusicology 33(2):217–227
7. Tenzer M (2000) Gamelan gong kebyar. The art of twenthieth-century Balinese music. University of Chicago Press, Chicago
8. Kunst J (1973) Music in Java, 3rd edn. Martinus Hijoff, The Hague
9. Helmholtz H (1863) Die Lehre der Tonempfindungen als physiologische Grundlage für die Theorie der Musik. Vieweg, Braunschweig
10. Nederveen CJ (1969) Acoustical aspects of woodwind instruments. Frits Knuf, Amsterdam
11. Benade A (1990) Fundamentals of musical acoustics, 2nd edn. Dover Publications, New York
12. Bader R (2013) Nonlinearities and synchronization in musical acoustics and music psychology. Springer series current research in systematic musicology, vol 2. Springer, Heidelberg
13. Bader R (2009) 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. Peter Lang Verlag, Frankfurt a.M, pp 95–112
14. Savage PE, Brown S, Sakai E, Currie TE (2015) Statistical universals reveal the structure and function of human music. www.pnas.org/cgi/doi/10.1073/pnas.1414495112
15. Tallotte W (2015) Meaningful Adjustments: music performance and ritual action in a south Indian temple. Anal Approaches World Music 4(1):1–22
16. Fletcher N, Rossing Th D (2000) Phyics of musical instruments. Springer
17. Hughes DW (1992) Thai music in Java, Javanese music in Thailand. Two case studies. British J Ethnomusicol 1:17–30

18. Miller T (2010) Appropriating the exotic: Thai music and the adoptation of Chinese elements. Asian Music 41(2):113–148
19. Jairazbhoy N (1971) The rāgs of North Indian music. Their structure and evolution. Faber and Faber, London
20. Rowell L (2000) Scale and mode in the music of the early tamils of south India. Music Theor Spectr 22(2):135–156
21. Minder A (ed) (1998) The Saṅgītopaniṣat-sāroddhāraḥ. A fourteenth-century text on music from western India. Composed by Vācanācārya-Śrī Sudhākalaśa. Indira Gandhi National Center for the Arts, New Delhi
22. Pondus TH (1974) Bagpipes and tunings. Detroit monographs in musicology. Nr. 3, Detroit
23. Harris I (2007) Buddhism under Pol Pot. Documentation Series No. 13. Documentation Center of Cambodia, Phnom Penh
24. Bizot F (2004) The gate. Vintage
25. Sam S-A (2008) The Khmer people of Cambodia. In: Miller TE, Williams S (eds) The garland handbook of southeast asian music. Routledge, New York, London, pp 89–102
26. Bader R (2011) Buddhism, animism, and entertainment in cambodian melismatic chanting smot. In: Schneider A, von Ruschkowski A (eds) Hamburg yearbook of musicology, vol 28, pp 283–305. Peter Lang Verlag, Frankfurt a.M
27. Naing M (2000) National ethnic groups of Myanmar. Swift Winds Books, Yangon
28. Sadam M (2013) Begin and becoming Kachin. Histories beyond the state in the borderworlds of Burma. Oxford University Press, Oxford, pp 242–253
29. Hertz HF (1902) A practical Handbook of the Kachin or Chingpaw language containing the grammatical principles and peculiarities of the language, colloquial exercises and a vocabulary with an appendix on Kachin customs, law and religion. Rangoon
30. Hanson O (2012) The kachins. Their customs and traditions. Reprint of 1913. Cambridge University Press, Cambridge
31. Lundström H, Tayanin D (1981) Kammu gongs and drums II: the long wooden drum and other drums. Asian Folklore Stud 40(2):173–189
32. Obayashi T (1966) The wooden slit drum of the Wa in the Sino-Burman border area. Beiträge zur Japanologie 3(2):72–88
33. Lundström H, Tayanin D (1981) Kammu gongs and drums I: the kettlegong, gongs, and cymbals. Asian Folklore Stud 40(1):65–86
34. Uchida R, Catlin A (2008) Music of Upland Minorities in Burma, Laos, and Thailand. In: Miller TE, Williams S (eds) The Garland handbook of Southeast Asian Music. Routledge, New York and London, pp 303–316
35. MacLachlan H (2011) Burma's pop music industry. Creators, Distributors, Censors. University of Rochester Press, Rochester, NY
36. Khin Myo Chit (1995) Colorful Myanmar, 3rd edn. Khin Myo Chit, Yangon
37. Roy M (2008) Cuban music: from son and rumba to the Buena Vista Social Club and timba cubana. Markus Wiener Publishing Inc
38. Zaw K (1981) Burmese culture. General and particular. Ministry of Information, Rangoon
39. Rossing T (2001) Science of percussion instruments. World Scientific Publishing Corporation, New York
40. Schneider A, von Ruschkowski A, Bader R (2009) Klangliche Rauhigkeit, ihre Wahrnehmung und Messung. In: Bader R (ed) Musical acoustics, neurocognition and psychology of music hamburger Jahrbuch für Musikwissenschaft 25, Peter Lang, Frankfurt a.M, pp 101–144

John Blacking Revisited—Comparative Analysis of Venda Tshikone Dance (1958 and 2009)

Jukka Louhivuori

One of the first research topics of interest for folk music researchers was the stability/instability of folk tunes. Particularly the question which parts and elements in folk tunes are sensitive to change and what are the stable elements of the melodies. Researchers had data collected over several decades, which gave a great opportunity to systematically explore stability of melodies over the years and decades. Some key principles were found, such as the fact that the endings of phrases are more likely to remain untouched compared to other parts of melodies [3, 6–8]. Even in the 1970s, the topic was still studied quite a lot [9, 11–14], but over the last few years the subject has been overshadowed by other themes.

New research methods and tools provide researchers today better tools for studying the question of change in music. Automatic notation methods make it easier for scientists to work and also allow them to analyse big data. Today, scientists have even better opportunities to deepen their understanding and knowledge of stability/instability of music.

John Blacking has been working in many areas of music research, but his research on Venda culture has been particularly influential. The book "How Musical is Man?" is based on field studies conducted in South Africa among Venda people in 1950s [1]. In this book Blacking gave convincing evidence of the limitations of how at that time musicality was understood. Social and musical structures related to Venda music appeared to be as complex as in Western classical music. In addition to cultural and social origin of musicality, his thoughts about the importance of understanding music related cognitive processes had an important impact on the later development of cognitive musicology.

J. Louhivuori (✉)
Department of Music, Art and Culture Studies, University of Jyväskylä, Jyväskylä, Finland
e-mail: jukka.i.louhivuori@jyu.fi

R. Bader (ed.), *Computational Phonogram Archiving*, Current Research in Systematic Musicology 5, https://doi.org/10.1007/978-3-030-02695-0_4

The material Blacking collected includes notes, photos, audio recordings, films and instruments. After Blacking's death in 1990, his wife donated Blacking's material to Professor John Callaway at the University of South West, Australia [10]. Some of the material was given to the School of Music at the University of Washington, USA (https://music.washington.edu/john-blacking-venda-music). The audio recordings in the present study have been taken from the University of Washington collection.

The purpose of this article is to find out how one of the most important dance among the Venda people, Tshikona, has changed between the years 1958 and 2009 (see [2]). The author of this paper had a possibility to collect Venda music and dance in the same areas where John Blacking made his ground breaking research in 1950s. The present study focuses on Thsikona dance, but in addition to it, video recordings were made also from Tshigombela dance.

Venda music has not been much studied. In addition to Blacking an interesting article has been written by Tracey and Gumboreshumga [15]. Kirby has written one of the first descriptions of Venda reed-pipe music [5].

1 Material

The study compares two Tshikona dance recordings. The first recording was made by John Blacking in 1958 in Shakadza village (https://music.washington.edu/john-blacking-venda-music; track 22) and the second in 2009 in the Gokolo village. During the same trip researchers visited two other villages, Mubvumoni and Tshidzivhe.[1] The data collected during this visit is stored in the archives of African music at the University of Jyväskylä in the Musirods server. The idea of this server is to store music related data in a format which makes it easy to store the data later into iRods or similar data managing system [4]. Musirods server contains at the moment data related to music therapy, music education and ethnomusicology. Venda material stored into Musirods was recorded in villages which are quite close to the Shakadza village, where Blacking made his recordings. The distance from Tshidzivhe to Shakadza village is about 60 km.

2 Dance Under Study

Tshikona is danced by men only. The dancers make up a large circle with a drum group in the middle. Each dancer has a pipe which gives one tone. Some of the pipes give two different pitches. Each dancer produces a pitch at a given time so that the

[1] In addition to the author of this article the following researchers attended in the field trip: prof. Jane Davidson (University of Western Australia), Dr. Edward Lebaka (University of Pretoria) and Dr. Pekka Toivanen (University of Jyväskylä). Thivhafuni Daniel Tshishonge (University of Tohoyando) was the guide during the visit.

Fig. 1 Dancers form a large circle with a drum group in the middle

desired musical product is created. Each dancer performs dance steps while playing the pipe (Fig. 1).

The pipes are sometimes referred to as reed pipes, even though most of the instruments are currently made of wood, metal or plastic. Getting a sound from the pipes requires very strong blowing. Especially from longer pipes it is hard to get a sound. Shorter pipes are lighter to play. The simultaneous blowing of tens, sometimes up to almost hundreds of pipes, produces a very powerful sound, close to big cathedral organs.

Today the instruments used are (1) wooden pipes, which are called vhutilo that produce two notes, (2) bamboo species called Musununu (reed pipes), that can only found in Tshaulu village around the whole South Africa, (3) plastic pipes (white and brown for electricity), which are used for the lower notes, and (4) metal pipes, which are also used for the lower notes (Fig. 2).

Fig. 2 A group of dancers were asked to play the pipe one after another. Thus, the pitch of individual pipes was possible to be measured accurately

The names of individual pipes from the highest to the lowest are: (1) Thakhula (do'), (2) Phala (ti), (3) Nzhingi (la), (4) Tshiaravhe (so), (5) Dangwe (fa), (6) Phalana (mi), ja (7) Kholomo (re) (Fig. 3).

Pitches of the pipes form the heptatonic scale f#, e, d#, c#, b, a, g#. Pipes in different octaves are used. Pitches of individual pipes are not tuned precisely to the same pitch, which gives for the sound a very special character (Fig. 4).

The drums are positioned in the middle of the dancers' circle. The drums are referred to as Murunzi, which means shade. Tshikona is supposed to be rehearsed under the tree which gives shade. Murunzi is the biggest drum used in Tshikona and the one that follows the steps of dancers and thus plays more varied rhythmic patterns. Thungwa is smaller than Murunzi and is the one which maintains the steadiness of the beat. Murumba is conical in shape and combines the two drums. Some Tshikona groups prefer only one while others can have more than one Murumbas (Fig. 5).

Fig. 3 Different sizes of Tshikona pipes

Fig. 4 The dancers are dressed with colorful cloaks, in some cases the cloak was white

Fig. 5 The drum group consists of two large drums (Murunzi and Thungwa), and two smaller ones (Murumba)

The material recorded during the field trip is stored in the digital archive of the University of Jyväskylä (Musirods). The material is categorised using the MARC 21 format. Table 1 is an example of the variables included in the video recording of Tshikona dance in this article.

Table 1 An example of the variables used in the classification of the data collected from Venda people (MARC 21 format)

	Field code	Subfield code
Date/time and place of an event (R)	033	‡a 20090217 ‡p Gokolo; Coordinates from Siambe Primary School 22° 56′ 6.40″ S, 30° 31′ 15.701″ E
Language code (R)	041	‡a Ven
Title statement (NR)	245	‡a Venda_Gokolo_Reedpipe1_20090217.mp4 (/.mov)
Media type (R)	337	‡a Video
General note (R)	500	‡a Traditional Tshikona reedpipe dance; dancers in a circulating ring formation
General note (R)	500	‡a Every dancer (male) has a reedpipe whistle; the drummers (female) don't dance
General note (R)	500	‡a Small children are mimicing the dance outside the circle
Formatted contents note (R)	505	‡a Instruments: reedpipes (major scale) Thakhula (8), Phala (7), Nzhingi (6), Tshiarache (5), Dangwe (4), Phalana (3), Kholomo (2)

(continued)

Table 1 (continued)

	Field code	Subfield code
Formatted contents note (R)	505	‡a Instruments: Drums Muruzi (biggest drum, follows the steps of the dancers), Thungwa (Maintains the beat), Murumba (Combines the two previous drums)
Subject added entry—personal name (R)	600	‡a Surname, first name ‡e funktion
Subject added entry—corporate name (R)	610	‡a Venda
Subject added entry—topical term (R)	650	‡a Traditional music, drums, traditional dresses, aerophones
Subject added entry—geographic name (R)	651	‡a South Africa, Limpopo, Gokolo
Index term—uncontrolled (R)	653	
Added entry—personal name (R)	700	‡a Louhivuori, Jukka ‡e cng
Added entry—corporate name (R)	710	‡a A Research Group from Jyväskylä

3 Methodology

Pitches were analysed using Transcribe! software. Blacking recorded Tshikona dances by staying in one position outside the circle of dancers. Because the dancers move in a circle one pipe in time will appear in front of the tape recorder's microphone. Thus, the sound of every single pipe was possible to be recorded. The sound of pipes, which were played next to each other in the circle, causes problems for sound analysis We used the same recording method as Blacking used. To get better audio recording for analytical purposes we recorded seven pipes, which form the heptatonic scale, separately. Tempo was measured by using a metronome. The metronome was adjusted to match the tempo of the dance. This simple method was possible to use because the tempo of the dance stays stable throughout the performance.

In addition to the pitches of pipes, changes in melodic and rhythmic structures between performances in 1958 and 2009 were analysed. Transcription was made by Olli Moilanen (see Appendix 1). The purpose was not to get an accurate transcription of the music, but to give an overview of Tshikona's melodic and rhythmic structure. Sibelius software did not give good possibilities to transcribe rhythmic micro level details.

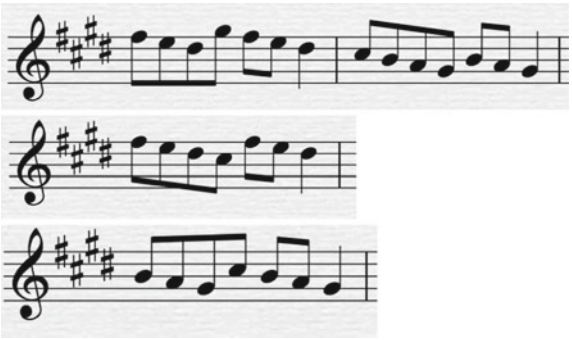
4 Results

In both recordings the same instruments were used: pipes and drums. Tempo in the 2009 recording appeared to be somewhat slower than in the 1958 recording. In 1958 the tempo was 101.3 bpm and in 2009 89.9 bpm.[2] In addition to Gokolo village, Tshikona was videotaped also in the Tshidzivhe village. In the performance of the dance group in this village the tempo was 92.0 bpm. Thus, tempo in both 2009 Thsikona performances was closer to each other than to the recording of John Blacking in 1958.

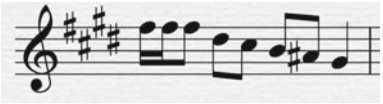
Human physiology sets certain limits to the tempos of dances. This may be one of the reasons why tempo is not as sensitive to changes as other aspects of performances. Comparative data in this study is too limited to make far-reaching conclusions about differences found in tempo between 1958 and 2009 performances.

In both recordings, the pipes are tuned to the same heptatonic scale using the following pitches: f# 3, e# 3, d# 3, c# 3, h 2, a 2, and g# 2.

Thsikona 1958 and 2009



Tshikona 2009; phrase which did not appear in 1958 performance.



Example 1. Basic melodic structure in Tshikona performances in 1958 and 2009 and the new phrase appeared in the 2009 performance.

[2]It has not been possible to check the details of the audio recordings by Blacking. The model of tape recorder is not known by the author of this paper. Thus, it is possible that tempo changes are due the equipment. Because the pitches of pipes are very similar, it is probable that Blacking's audio recordings give quite reliable picture of tempos of dances and frequencies of pipes.

Table 2 Pitches of Tshikone pipes in 1958 and 2009 recordings

Name	Solfa	Pitch	1958	2009
			Hz	Hz
Thakhula	do'	f# 3	188	188
Phala	ti	e 3	167	170
Nzhingi	la	d# 3	157	158
Tshiaravhe	so	c# 3	140	146
Dangwe	fa	h 2	125	128
Phalana	mi	a 2	116	114
Kholomo	re	g# 2	107	110

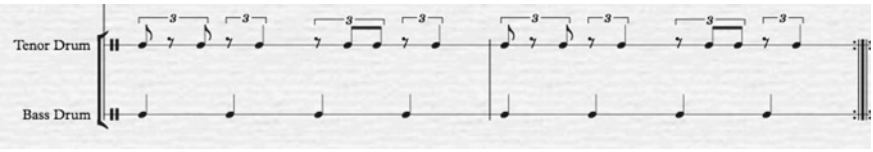
In the 1958 recording the melodic structure of Tshikona did not contain any other major melodic elements than those transcribed in the transcription above (Example 1). In the Tshikona version recorded in 2009 a new melodic pattern was performed by high pitched pipes. This melody imitates the song sometimes sang during Tshikona performances. In one of the 1958 recordings a male singer sings a phrase, which resembles the phrase the high pipes played in 2009. In the Thsikona dance video recorded in 2009 in Tshishive village a female singer sang this phrase in the beginning of the performance. The phrase played with high pitched pipes had a# instead of a, thus, at the same time two scales were played simultaneously (bitonality).

The tuning of pipes has remained almost unchanged. In both recordings the highest tone of the scale f# 3 has a frequency of 188 Hz. The biggest difference is in the c# 3, which was in 1958 at 140 Hz and in 2009 146 Hz. However, it is important to notice that the pitches of individual pipes vary and thus the comparison made in the Table 2 is an example of a few individual pipes.

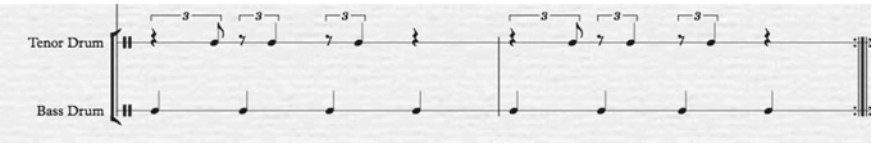
In the Table 2 the pitches of pipes are analysed from the 1958 and 2009 recordings.

The basic rhythmic pattern in 1958 and 2009 is pretty similar representing the so called African pattern:

Tshikona basic rhythmic pattern in 1958



Tshikona basic rhythmic pattern in 2009



5 Conclusions

Comparison of recordings made in 1958 and 2009 shows that Tshikona dance has preserved many of its basic features: the pitches of the pipes are the same and the tune is based on the same heptatonic scale. The tempo has somewhat slowed down. Also the basic rhythmic patterns played by drums are the same in 1958 and 2009. Based on the analyses of just a few Tshikona performances, it is not possible to say with certainty whether differences found are representations of typical local variations or examples of more persistent changes happened over the last fifty years. Analysed Tshikona recordings support interpretation that Tshikona dance was performed in 2009 in a style which is pretty close to the tradition of earlier decades. Not much changes in key elements of performance were found in comparison of the recordings from 1958 and 2009.

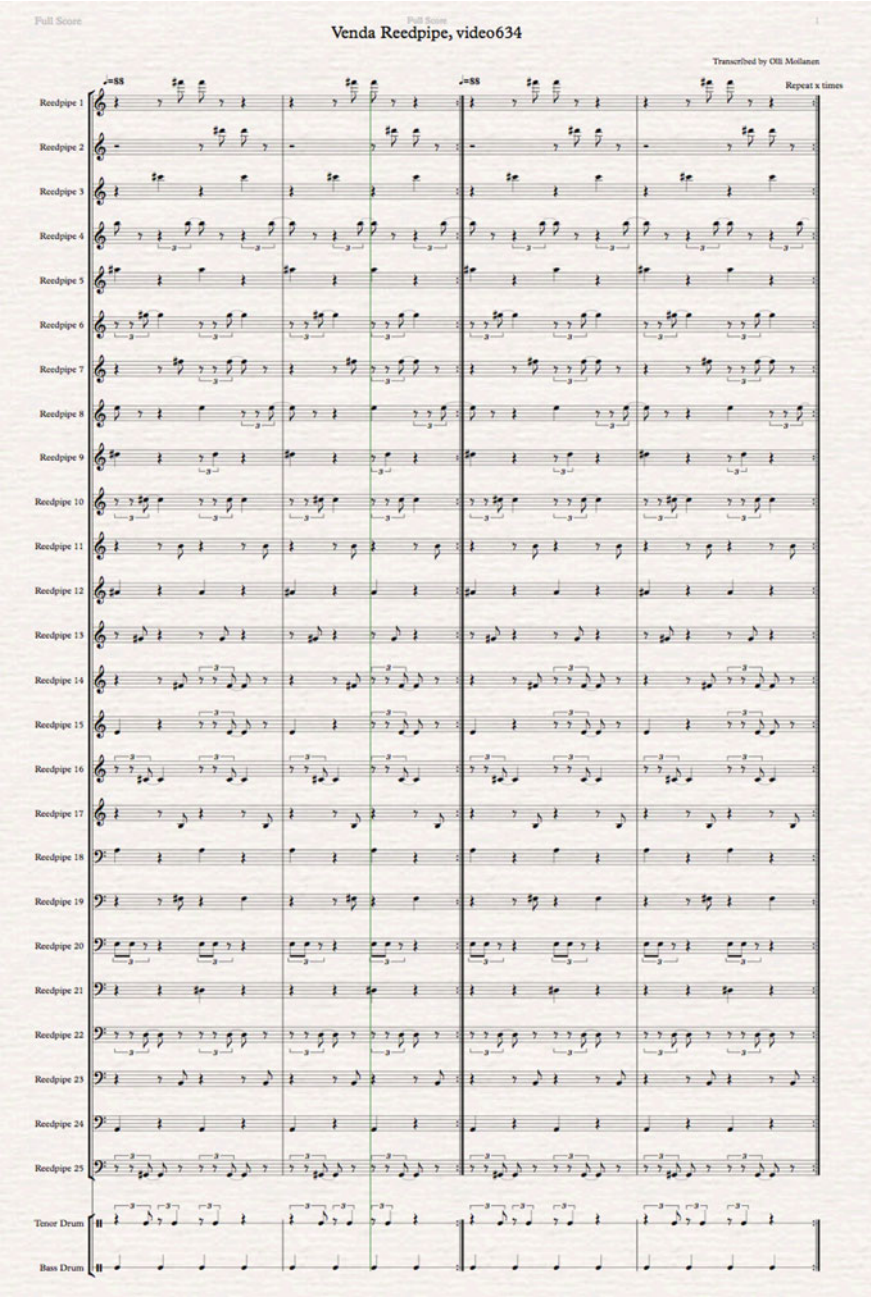
During the field trip in 2009 video recordings were made also from a few Tshigombela dances. This material is not yet fully analysed, but preliminary comparison between Tshigombela performances in 1956 and 2009 has been done. The film and audio recordings done by John Blacking from Tshigombela dance shows that performances of this dance are pretty similar in 1958 and 2009: tempo, dance steps, instruments used, melodic and rhythmic structures and even dressing is not much changed during decades.

In the future more careful analyses of data in Blacking and Jyväskylä collections should be made to get more reliable information about how music is changed in Venda culture during fifty years period.

Acknowledgements I want to thank Dr. Morakeng Edward Kenneth Lebaka of his very valuable help in organizing the fieldwork in Venda. His experience in field work and knowledge about African culture and music had a great value for the success of this project. Also I want to thank Thivhafuni Daniel Tshishonge who guided us during the field trip. His deep knowledge about Venda culture and connections to Venda people make it possible to get connections to Venda people and visit the villages.

Appendix 1

Transcription of the Tshikona dance in the Gokolo village in 2009 by Olli Moilanen.



References

1. Blacking J (1973) How musical is man? University of Washington Press, Seattle, USA
2. Blacking J (1977) Some problems of theory and method in the study of musical change. Yearbook Int Folk Music Council 9:1–26
3. Bronson BH (1951) Melodic stability in oral transmission. J Int Folk Music Council 3:50–55
4. iRods (2018). https://irods.org. Accessed 29 June 2018
5. Kirby R (1933) The reed-flute ensembles of South Africa: a study in South African native music. J R Anthropol Inst Great Br Irel 63:313–388
6. Koller O (1902) Die bäste Methode, Volks-, und volksmässige Lieder nach ihrer melodischen Beschaffenheit lexikalisch zu ordnen? Sammelbände der International Musikgesellschaft IV 1–15
7. Krohn I (1899) Über die Art und Entstehung der geistlichen Volksmelodien in Finnland. Helsinki
8. Krohn I (1903) Welche ist die beste Methode, um Volks- und volksmäßige Lieder nach ihrer melodischen (nicht textlichen) Beschaffenheit lexikalisch zu ordnen? Sammelbände der Internationalen Musikgesellschaft 4:643–660
9. Lindblom B, Sundberg J (1970) Towards a generative theory of melody. Svensk Tidskrift för Musikforskning 52:71–88
10. Malacari G, Campbell PS (2003) The work and legacy of John Blacking. Dvd disc. Callaway Centre, School of Music, The University of Western Australia
11. Merriam AP (1955) The use of music in the study of a problem of acculturation. Am Anthropol 57:28–34
12. Nettl B (1964) Theory and method in ethnomusicology. Free Press of Glencoe, New York
13. Nettl B (2005) The study of ethnomusicology, 2nd edn. University of Illinois Press, Urbana
14. Sundberg J (1975) On melodical similarity in versions of a Swedish folk-tune. STL-QPSR 16(2–3):61–66
15. Tracey A, Gumboreshumba L (2013) Transcribing the Venda Tshikona reed pipe dance. Afr Music J Int Libr Afr Music 9(3):25–39. https://doi.org/10.21504/amj.v9i3.1909

Smithsonian Folkways and the Associated Ralph Rinzler Folklife Archives and Collections

Jeff Place

Abstract Smithsonian Folkways and the associated Ralph Rinzler Folklife Archives and Collections were created in 1988 as a vehicle to collect world music record labels and disseminate the recordings in a number of ways. In addition to the recordings of the 50 years of the Smithsonian Folklife Festival which has included participants from all over the world, the Smithsonian Institution has been collecting the recordings of independent record companies, 13 to date. This includes Folkways, Arhoolie, Paredon, Cook, the Mickey Hart World Music Collection and the UNESCO labels. The plan of collecting record labels allows us to gain all the intellectual property rights. This assures we can legally distribute the recordings and see that the musicians, informants, and communities receive their due royalties. The recordings as distributed as commercial CDs or LPs, streaming audio, or downloads. One can download tracks through our website or services like Itunes. We have created over 400 new recordings since 1988 and have over 4000 recordings in our catalog. Our robotic Micro-Tech machine can create compact discs of the 4000 titles. The roughly 4400 commercial recordings are about 15% of what exists in the Rinzler Archives. The rest consists of festival recordings and outtakes from the recording labels. This paper hopes to share a brief glimpse into our processes.

The parallel founding of Smithsonian Folkways Records and the Folklife Archives for the then Smithsonian Office of Folklife Programs in 1988 was the beginning an interesting experiment. It was created as a hybrid archive of world music and sound and an active record label residing in the national museum of the United States. Prior to that time the archive had been the materials from the Smithsonian Folklife Festival (1967-present) and other activities of the Folklife Office. Folkways Records, itself, had existed as an independent record company in New York City, putting out 2168 albums of the sounds of the twentieth-century.

Ralph Rinzler (1934–1994) had grown up a fan of opera but joined the mass conversion by many to a fan to folk music which occurred in the United States in the

J. Place (✉)
Smithsonian Institution, Washington, D. C., USA
e-mail: PlaceJ1@si.edu

R. Bader (ed.), *Computational Phonogram Archiving*, Current Research in Systematic Musicology 5, https://doi.org/10.1007/978-3-030-02695-0_5

1950s. He worked as a musician with the Greenbriar Boys in New York recording for Vanguard Records. While in college at Swarthmore College in Philadelphia he became very involved with programming of folk music events. He brought that skill with him to New York. Rinzler, like many, had been influenced by the 1952 Folkways release, Harry Smith's important *The Anthology of American Folk Music*. The Anthology had put back into print 84 important 78 rpm recordings of American folk, Cajun, blues and gospel music from 1926–1934. It became the repertoire of many of the 1950s folk musicians in the United States and elsewhere. Rinzler and other important movers and shavers in the folk revival realized that these recordings were only 25–30 years old and that many of the musicians were likely still around. They travelled south and located many, bringing them north of folk festivals. Rinzler had a key role at the Newport Folk Festival, and during this time, he brought recordings of his re-discoveries to Folkways Records, the label at the time most likely to issue them. Through this he established a relationship with Folkways' owner Moses Asch. This would prove important later.

Rinzler's career later took him to the Smithsonian Institution in 1967 as the first artistic director of the new Festival of American Folklife. He later founded the Smithsonian Folklife Program in 1977. By the mid-1980s he had risen to Smithsonian Assistant Secretary for Public Service. His job was to oversee all of the parts of the Smithsonian involved in public outreach.

Moses Asch (1905–1986) was the son of novelist Shalom Asch. The family emigrated to New York in 1915 from Paris. Asch as a young man became involved in radio engineering, opening up a shop in New York. Many of customers bemoaned the fact that they could not purchase ethnic Jewish 78s anywhere in the city. Asch entered the record business to fill this need, releasing a recording by the Bagelman Sisters in 1939. One of the famous stories about Asch at that time was that his father was asked by his friend, the scientist Albert Einstein, if he knew of someone who had a recording device. Einstein wished to record a message for the Jewish people in Germany to leave as soon as possible. Sholem Asch mentioned his son and they travelled to Princeton, New Jersey. At dinner Einstein inquired of the younger Asch, "Well, Mr. Asch what do you do?". He replied "I repair radios and install public address systems but my dream is to create a full encyclopedia of the sounds of the Earth". Einstein replied in vigorous support.

Asch followed through first with his Asch and Disc labels and then finally the main event, his Folkways label, which started on May 1, 1948. Over the course of the next 39 years he created his over 2000 albums. It included folk music by such figures as Woody Guthrie, Lead Belly and Pete Seeger. It was also strong on children's music, spoken word and sound effects, both manmade and natural. At the time of Folkways founding there were not many record companies who would release ethnographic field recordings from around the world as well as immigrant groups in the United States. Starting with his Disc label and continuing with the Ethnic Folkways Library under editor and anthropologist, Harold Courlander they released hundreds of titles from around the world. He spent a good deal of time going to academic conventions and speaking to professors who found Asch to be the place they could go to get their work published. Asch also felt strongly about including complete liner note booklets

with each release. His main customer base was schools, universities and libraries. By the 1960s Asch had created a full map of the world showing where all his recordings were made. According to his son, Michael, an anthropologist and former professor at University of Alberta, it was Moses Asch's goal to fill in the whole map. He was always looking for those missing pieces.

In 1986, Moses Asch at Folkways was thinking of retiring. He was slowing down and was taking offers for his label. Rinzler, who had a fondness for the label, contacted Asch and the Smithsonian threw its hat into the ring. Initially Asch was hesitant. He had a long history of working with musicians and writers were affected by the government harassment and the blacklisting in the 1950s. Having his life's work associated with the United States Government made him fear for what effect they could have on his encyclopedia. He often had released records with very little sales potential but he felt for a democracy to fully function it need an educated populace, there was a "need to know". He even officially called his label, Folkways Records and Service Corporation. The Smithsonian agreed to Asch's stipulation that whomever took over Folkways had to keep every single title available forever even if it sold one copy every five years. No profit minded company would agree. Rinzler meanwhile was trying to counter the conservative curators at the Smithsonian who argued that the national museum had no place collecting sound recordings, the museum was about objects, science and art. A deal was reached in 1987, between the Asch estate the Smithsonian and the collection came to the Smithsonian in fall of 1987.

As part of the negotiations, the Smithsonian agreed to fund two positions, one a curator and director to oversee the continuation of the label and the collection. The second, an archivist to oversee the archival collection. Chosen as director/curator was ethnomusicologist Anthony Seeger. I was chosen as archivist. Our goal was to continue Folkways and it also led to the start another record label, Smithsonian Folkways to carry on Asch's life's work with new projects in the spirit of Folkways. We also started to acquire other independent record labels similar to Folkways, where the label's creator wanted their legacy to endure.

For the almost four decades of Folkways, Asch had operated a small shop. He did all of the production work and his partner, Marian Distler handled day to day operations in the office. He worked long hours and managed to issue the equivalent of a record a week for 40 years. Granted, some of the years of the 1960s saw more releases than the 1970s or 1980s. Now at the Smithsonian, the label had to work in a giant bureaucracy-forms and extended periods to get paperwork through. In the early years, Anthony Seeger was able to get much done through having our record distributors handle much of the business and production. Eventually all was brought into the Smithsonian office. Even with these ongoing difficulties Smithsonian Folkways has been able to release 400 recordings over 30 years, not quite up to Moses Asch's pace,

The archive was set up as a partner to the label. There is an issue with running a business with U.S. federal funds, even if the label is non-profit, which it is. The Smithsonian has what it calls "trust funds". This is based on endowments starting with $500,000 given to the institution in 1835 by the estate of James Smithson to found under the name of the Smithsonian Institution an establishment "for the

increase and diffusion of knowledge among men". All of the activities of Smithsonian Folkways are run with trust funds, philanthropic gifts, grants, or proceeds of the sale of recordings. The archive is a different matter, support for the archive is federally funded, it a federally owned collection. This supports staff and the maintenance of the collection.

The collection, itself, is the paper files of Folkways, the production files and album layout materials, and the audio recordings, masters and outtakes. During the 1990s, work was done with a small staff to catalog, identify, and to migrate many of the recordings to a more stable form. Unfortunately, some of the sound was copied to digital audio tapes so needed to be redone. The collection has 5000 instantaneous discs, some of which are ethnographic field recordings, some are important recordings made on glass acetates in Moses Asch's studio. There are about 10,000 open reel tapes, the earliest are Fred Ramsey's recordings of the singer Lead Belly in 1948. About half are masters for the over 2000 albums. The other half are out takes, the full set of field tapes or duplicate safety copies. Extensive inventories of these recordings have been done, identifying content on unmarked or poorly labelled recordings. Preserving them for the future is an important goal, but more importantly the mission of the archive, as with Smithsonian Folkways, has been outreach, to see them used and made available to the public in a myriad of ways. Many of the 400 recordings released by Smithsonian Folkways since 1988 have come from the archive, including many unreleased recordings discovered during the cataloging and transfer process.

Meanwhile starting in 1988, decisions were being made as to what we should publish in the new compact disc era. Anthony Seeger had a group of ethnomusicologists in various specialties access the quality of the recordings previously issued on Folkways. Certain historically important works were identified. He prioritized recordings where the traditions had either ceased to be performed or had changed dramatically. He felt if the tradition still was much like before it was better to document it now using better equipment. Recordings like John Cohen's Peruvian recordings or Colin Turnbull's Ituri Pygmy recordings joined the compact disc era,

We started actively seeking other extinct or soon to be record companies. The second collection what that of recording engineer Emory Cook. The Cook or Cook Labs label was active in the 1950s and 1960s. Cook was an inventor of high end audio equipment and state of the art production processes at the time. His records were prized by individuals with state of the art audio gear for their stellar fidelity. Cook approached us in 1990 about whether we wished to acquire his label. We don't have an acquisitions budget to buy labels, so this started our policy of having the collection appraised so the donor can get a tax deduction, have it a straight donation, or to find a philanthropist to help fund the acquisition.

In subsequent years, we have acquired other labels. Why labels? It goes to a core philosophy of what we do. Acquiring the label and business papers allows us to acquire the intellectual property rights and contracts to the material. We can then use them freely for our publications, license them to third parties, create new versions or unreleased projects. Very importantly having the contracts allows us to pay the creators for their work. In the case of those creators we can not locate we devote a good amount of time trying to find them. This often means spending hours and hours

of staff time to find someone to mail a twenty dollar check. In some cases, we have arranged to send the money to a common fund within a community

We are often approached by individuals wishing to donate collections. In many cases, and it was common practice when the recordings were made, the donor has no identification of any of the musicians on the recordings. The royalty recipient was thought to be the collector or archive. For us to ethically publish them we would need to track down the original musicians or estates which makes it untenable.

After the Cook Records acquisition we have subsequently acquired others. In 1990, we acquired Paredon Records from Barbara Dane and Irwin Silber, a label of music from political movements worldwide. We acquired singer Richard Dyer-Bennet's label. Fast Folk Records was a singer-songwriter collective in New York which put out over 100 titles of emerging singer-songwriters. Monitor Records, founded by Rose Rubin and Michael Stillman, of music from around the world, including strengths in Eastern Europe and the former Soviet Union. Others were Collector Records (trade union songs from the U.S. and U.K.), MORE Records (mariachi music of New Mexico); BRI Records (Appalachian music from Virginia); The London Library of Recorded English (spoken word of British literature); the recordings of world music made by and issued by Grateful Dead percussionist Mickey Hart.

Important recent acquisitions are the UNESCO record label. These recordings had been issued in various ways over the years most recently by the Audivis label. It includes about 150 titles, including ones that not been issued that we subsequently have issued. We also acquired the Arhoolie label, run by Chris Strachwitz from 1960–2016. Arhoolie includes approximately 500 titles of American and Mexican vernacular music including norteno, mariachi, zydeco, blues, Cajun, and gospel music. It also includes the recordings by the Ideal, Falcon and Discos Smith labels. Arhoolie also released anthologies of older 78 recordings from around the world.

Each one of these collections can include the papers of the founders, the labels masters and outtakes. Also, it could commercial records or books that were the property of the former owner, that are added to our reference library. These are not distributed by us, just library researchers can use. We have also acquired other recordings by recordists where these are the remaining recordings that complement issued titles on our distributed labels. Examples of this would be the recordings of Charles Bogert and Arthur Greenhall (animal sounds) and musical recordings by Verna Gillis, Ralph Rinzler, Eric Davidson, Tom Wisner and Frederic Ramsey Jr.

In the early 1990s, we received a grant from the Ford Foundation to create Smithsonian Global Sound. The idea was to create a portal and an audio "union catalog" of recordings from ethnomusicology archives around the world in every country. It would be "one stop shopping" for researchers in world music and a resource for teachers (Fig. 1). ARCE (India) and ILAM (South Africa) joined in. There were discussions with others but the idea did not take off. There has been discussion about reviving it (Fig. 2).

The archive continues to look for other record company collections, looking to fill in gaps and to be able to tell a wider story. It also houses the materials created by our larger office the Smithsonian Center for Folklife and Cultural Heritage. The main

Fig. 1 The Microtech robotic duplication machine for public requests for compact discs

collection here is the materials of the 50 years of the Smithsonian Folklife Festival. Each year an average of two or three themes are chosen. These can be countries, an U.S. state, a region, or a theme based on occupational folklore. A typical year might

Fig. 2 Archive staff using the WaveLab software to digitize reel to reel tapes

be 2003 where we featured Mali, Scotland and Appalachia. 2002 was one program tracing the Silk Road from Japan to Italy. 2017 is Armenia and Catalonia. Frequently fieldwork is done before hand with audio recordings of oral history and music. In keeping with desire to own the rights for use we get releases signed for non-profit educational and museum use, allowing use.

All in all, these are the holdings of what is now called the Ralph Rinzler Folklife Archives and Collections, named after the founder of the office. The archive's goal is to manage and protect these collections and make them useable and available for the Center and Smithsonian Folkways. The collection is used to create on-line education plans. Frequently teachers will add their plans to the Smithsonian Folkways website.

The 50,000 tracks in our collection where we own full commercial rights are available through Smithsonian Folkways. They can be acquired by purchase on on-demand compact disc (or commercial jewel box CDs in the case of some). The on-demand CDs is truly a DIY (Distribute it yourself) system. We have two Microtech robotic system which can make copies for the asker. Once someone orders one, a message is sent to the robot, it copies the audio from a server to the disc, it also copies a.pdf of the liner note booklet to the disc. A sticker which is a copy of the original album jacket is printed out and affixed to a cardboard generic CD sleeve. It resembles a small LP. They can be shrink wrapped to sell in stores who wish them. Some titles are being reissued on vinyl while the demand remains.

The record business is recent years has seen the sales of physical recordings plummet. For a label/collection that has a goal to get our recordings and the story behind them out a new direction must be looked for. For much of popular music, a simple artist name and title might be adequate as what appears on the playback device. For us, the liner notes are crucial. If I look at my phone and it tells me the stream I just downloaded is "Deer Dance" but tells me nothing about which tribe, or how does it fit in their culture, that is not fulfilling our mission. All of the liner notes to all 4000 titles can be downloaded free from the Smithsonian Folkways website. Streaming is the way people are listening to music in the digital age and a stream pays a fraction of a cent to the creator. That is not sufficient to support an artists or a record company. We have had 250 million streams but the money coming in does not make up for what was lost of physical sales.

We have out materials on Itunes and other services but downloads are no longer selling like they were. We are looking for other ways. We are non-profit but need to break even. We have a service called Alexander Street Press through which schools, libraries and universities can subscribe to our entire catalog and students can have full access on campus. Our website takes certain parts of the collection and posts features online, some through our on-line *Smithsonian Folkways Magazine*. Outside scholars are used to create articles about their specialties.

As compact disc sales drop, box sets have held on so going the direction of a full box with recordings make sense, it fits our mission. Staff and fans create playlists we can post on-line and through social media. We have started a membership program, much like public television, which allows members access to our catalog and inside access to the staff and label. It is a different business model which will be necessary.

Back on the archival side changes are also happening. The titles available through Smithsonian Folkways make up only 15% of the recordings in the archive. What of all the historical papers and manuscripts? The paper files and media are being digitized through in-house Smithsonian grant funds. We also did receive a large outside grant called Saving America's Treasures for audio digitization. The paper materials and photographs are being scanned and imbedded metadata added. For audio we have been digitizing the recordings at 96 K, 24bit rates using the Wave Lab software. We are using BWF Metaedit to embed metadata in the audio files. For all of the digitized assets, they are being stored in the Smithsonian's Artesia Digital Asset Management System. The goal of the DAMS is to store assets but also to connect with on-line Smithsonian sites including the SOVA Smithsonian Collections search function. Assets can be made available to listen or view through this site depending on intellectual property rights or sensitivities with the creator or copyright. Certainly, the parts of the archive commercially available will not be posted free. Hopefully having the materials available on-line will allow researchers to help us with metadata on recordings where not all not all the metadata is known. The original recording, usually analog, will be maintained for future use if needed. We have witnessed three generations of "best practices" for migrating the sound over the last 30 years. There is no doubt there will be a fourth or fifth.

We hope we can raise funds in order to make all our commercial recordings ultimately available free to students all over the world through teachers. Making these materials and the many recordings in the archive, where we can, available on-line is exactly what we do and should do. It is using the label and on-line Smithsonian channels to get the content of our collections out there where individuals all over the world can use and learn from them. It is using the partnership between an internationally known non-profit record company and the wonderful collections in our archives to maximize their use to the best of our abilities for the people who seek to use them.

Analysis and Perception of Javanese *Gamelan* Tunings

Gerrit Wendt and Rolf Bader

Abstract *Gamelan* music as performed on the Indonesian islands of Java and Bali is one of the most well known and well studied non-European music traditions. Especially the tunings of *gamelan* ensembles have fascinated researchers since the early stages of the disciplines of Systematic and Comparative Musicology, as these tunings are considerably different from Western, Middle Eastern or other music traditions and even differ between *gamelan* ensembles. Additionally most instruments of the ensemble are percussion instruments with inharmonic overtone spectra. One way to explain certain characteristics of *gamelan* tunings is by relating the inharmonic sound spectra of the percussion instruments to the perception of dissonance. This theory is assuming that the psychoacoustic sensation described as auditory roughness is perceived as dissonant and is therefore avoided. This study investigates the influence of musical roughness on the perception of different *gamelan* tunings by correlating psychoacoustic measurements with a perception test in an online survey. Based on sound samples of an existing *gamelan* ensemble set based in Hamburg, Germany, a *gamelan* tune was built in a DAW. By detuning the sounds, several versions of the tune in different temperaments were built. It appears that the measured roughness perception and roughness measurements of the different tunes correlate very well for all detuned cases. Still the original piece does not correlate, pointing to a different perception strategy for the original ensemble tuning.

1 Introduction

The term *gamelan* refers to different ensemble types of the Indonesian Islands of Java and Bali. All of these ensembles use tuned percussion instruments, however the instruments as well as the music that is played on them can differ between different

G. Wendt (✉) · R. Bader
Institute for Systematic Musicology, Neue Rabenstrae 13, 20354 Hamburg, Germany
e-mail: gerritwendt@gmx.net

R. Bader
e-mail: R_Bader@t-online.de

R. Bader (ed.), *Computational Phonogram Archiving*, Current Research in Systematic Musicology 5, https://doi.org/10.1007/978-3-030-02695-0_6

musical traditions. The following paper is focusing on the Central Javanese *gamelan* ensemble in contrast to the two big neighboring traditions of Sunda (West Jawa) and Bali, as well as other smaller ones. This Central Javanese *gamelan* is often associated with the four big courts of Central Java and the two court cities of Yogyakarta and Surakarta. However, regional styles also exist within this tradition [1]. Sung poetic verses in old Javanese language (*tembang macapat*) are often a central part of the music, in contrast to other *gamelan* traditions. The umbrella term *karawitan* is used to encompass all aspect of the music.

1.1 Parts of a Central Javanese gamelan

Most parts of an *gamelan* ensemble consist of tuned percussion instruments. These can be divided by their shape into gongs and bar types. Preferably bronze is used as a material for their construction, but ensembles made out of brass or iron also exist. The frames of the instruments are made out of wood. The Gong instruments can be further divided into hanging gongs and lying gongs that have a wooden frame below them. Bar instruments can be divided into those that have tubular resonators and those that have a through-like wooden frame. The *gambang* is an exception because its bars are made out of wood. Besides tuned percussion instrument, different kind of drums (*kendhang*), a spike fiddle (*rebab*) as well as a bamboo flute (*suling*) are used. A male choir (*gerongan*) and a female solo singer (*sindhen*) represent the vocal parts of the ensemble.

1.2 Tuning Systems of Central Javanese gamelan *Music*

Central Javanese *gamelans* are tuned to either one of two distinct tuning systems that relate to two sets of tones (*laras*). *Slendro* is a pentatonic tuning characterized by intervals of almost the same size. *Pelog* is a heptatonic tuning, characterized by intervals of different sizes. Despite *pelog* being heptatonic as a tuning, different scales are formed in musical practice that are besides the use of additional tones pentatonic by nature [2].

An ensemble often consists of a double set with instruments of the same type in both tunings.

A system of musical modes (*pathet*) exist for each tuning. Each mode is characterized by its musical range, certain final, dominant and avoided tones, special melodic pattern (*cengkok*), that are realized by the elaborating instruments and a specific mood associated with it. This can partly be explained with their use in the Javanese shadow puppet theater *wayang kulit*, one of the main occasions where *gamelan* music is played. *Gamelans* are not tuned to a single standard, so that the intervals as well as the pitches vary between different ensembles.

1.3 Historical Development

The historical development of the musical instrument and the tuning systems is in many parts uncertain due to a lack of historical evidence. One question that arises when dealing with the historical development of *gamelan* ensembles and its music is, if the tuning systems and musicals instruments developed simultaneously or independently. The gong instruments are said to have been developed out of ancient bronze drums that were found in different parts of South East Asia [2, 3]. Old Javanese poems and depictions on temples indicate that gongs were in the past used as signal instruments [4, p. 20], on the other hand sung poetry (*tembang*) also using *pelog* and *slendro* scales is said to be the source for many modern gamelan pieces.

One source that might give further information about the historical development are old ceremonial ensembles that are owed by the courts of Java. They often differ from modern ensembles by consisting of fewer instruments, having a lower tuning in absolute pitch, being played instrumentally without singing and only on very rare occasions. The few surviving *gamelan munggang* are believed to have already been considered ancient in the 13th century [3]. They consist of gong instrument and their tuning only encompasses three tones (high, middle, low) that are often described as having a *slendro* or *pelog* character [3, p. 161, 181]. The *gamelan sekaten* are attributed to the 16th century. Few of them are found in the courts of Java and are only played once a year during the Islamic *sekaten* festival. They consist of bar and gong instruments like modern *gamelans*, but only the loud-sounding instruments are used and are tuned to a seven tone *pelog*, with instruments of bigger size and a lower pitch than modern ensembles [5, p. 161]. All those characteristics indicate that during the historical development of the modern Javanese *gamelan* more tones and instrument where successively added. The description and distinctions from the early 19th century by Stamford Raffles, who was governor of Java during a short British intervention, show that most of the musical instruments as well as many musical pieces of the time, are still in use today [6].

1.4 Theories About the Origin of Tuning of gamelans

The growing interest in non-European tunings and scales in the end of the 19th, beginning of the 20th century lead to an extended research on *gamelan* tuning, that was often based on comparative tone measurements. The ethnomusicologist Roger Vetter even assumes that the cent unit introduced by John Alexander Ellis "has been applied in the writings on Javanese music to a greater degree than in the literature on any other single tradition" [7, p. 217]. Many of these theories have a rather speculative character and where almost entirely based on the measurements of the fundamental frequency. One of the most represented is the so called theory of overblown fifth,

formulated by Erich Moritz von Hornbostel, which postulates that *pelog* and *slendro* scales can be derived out of a cycle of overblown fifth produced on closed pipes [8].

1.4.1 Auditory Roughness and *gamelan* Tunings

A more recent approach by Sethares [9] is not only limited to the fundamental frequencies but is relating the partials of the musical instrument to the tuning of the *gamelan*. Sethares' theory is related to the observations of the 19th century German physician Hermann von Helmholtz, who defined the buzzing sensation created by two close by frequencies as auditory roughness.

In the 20th century auditory roughness was explained with the limited frequency bandwidth of the auditory filters also known as critical band [10].

Based on the observations of coinciding violin tones, Helmholtz created a graph that depicts the auditory roughness of different intervals. Most intervals considered consonant in Western music are near the points of minimal auditory roughness [11, p. 318].

Similar graphs that portray the points of auditory roughness of the tones of single *gamelan* instruments are created by Sethares using statistics software. By using generic sound spectra of *bonangs* and *sarons* and combining them with harmonic partials, Sethares found that the minimal points of auditory roughness of the *bonang* gongs are close to the intervals of *slendro* and those of the metallophone *saron* lie near to the intervals of *pelog*. The tunings of the *gamelan* could therefore be explained as a result of a combination of spectrum with the minimal points of perceived auditory roughness [9].

Sethares' approach has many advantages compared to earlier theories. It is based on empirical results and on a psychoacoustic phenomenon, rather than on cultural factors.

However, only roughness curves of single musical instruments in combination with harmonic sound spectra were used while a *gamelan* ensemble consists of instruments with different structures of partials and therefore different minimal points of auditory roughness when played together at the same time. Furthermore, is it not clear if auditory roughness is cross culturally perceived as dissonant. In many musical cultures roughness is created purposely as an integral part of the music [12].

1.5 Research Questions and Overview

If *slendro* and *pelog* were the results of the combination of the spectra of the different parts of a *gamelan*, we would assume that *gamelans* which are tuned differently have a higher perceived level of auditory roughness. Assuming that roughness cross-culturally leads to the perception of dissonance, they should also be perceived as more dissonant. However, in a number of modern contexts [13–15] *gamelans* with non-traditional tunings are in use. In order to further study this approach, we sug-

gest to analyze the perception of auditory roughness and perceived dissonance of traditionally and non-traditionally tuned *gamelans*.

To do that, we decided to use a combination of measurements of musical roughness using signal processing software, as well as an online survey to test the subjective perception of tuning. By comparing the measured roughness values with the perception of different tunings, we hope to get a deeper insight into the role of auditory roughness in the perception of *gamelan* tunings. For the stimuli we decided to use *gamelan* samples of an existing ensemble that are arranged in a digital audio workstation (DAW) in the way to end in a classical *gamelan* piece. Using a musical pieces, rather than just sound samples of single instruments as done in previous research, is expected to provide a higher ecological validity.

2 Method

2.1 Stimuli

2.1.1 Recording and Sampling

The gongs and bars of the musical instruments of the *pelog* half of the *gamelan Margi Bodoyo* from the Indonesian Consulate in Hamburg were separately recorded. The frequencies of the recorded samples were measured. By comparing the measured frequencies with measurements done by Surjodiningrat et al. [16] it was concluded that the tuning shows similarities to a number of *gamelans* from Surakarta (Fig. 1).

In the next step, the recorded tones of a number of instruments that are sufficient enough to represent a small *gamelan* ensemble where chosen and sampled according

Comparison of tunings

Ensembles	Tones							
	1	2	3	4	5	6	7	i
Margi Bodoyo (Gender bem)	295	332	358	-	442	476	-	619
Margi Bodoyo (Demung)	303	332	357	426	451	475	525	-
Mardiswara	307	331	355	424	448	478	527	617
Kanjutmesem	295	320	347	406	440	470	519	598
Average Pelog	279	299	324	381	412	439	481	560

Fig. 1 Comparison of the fundamental frequencies in hertz of the forth octave register of the instruments *pelog demung* and *pelog bem gender barung* with the frequencies of *Mardiswara* and *Kanjutmesem* from Surakarta measured by Surjodiningrat et al. [16] as well as average *pelog* of all ensembles measured by the same authors

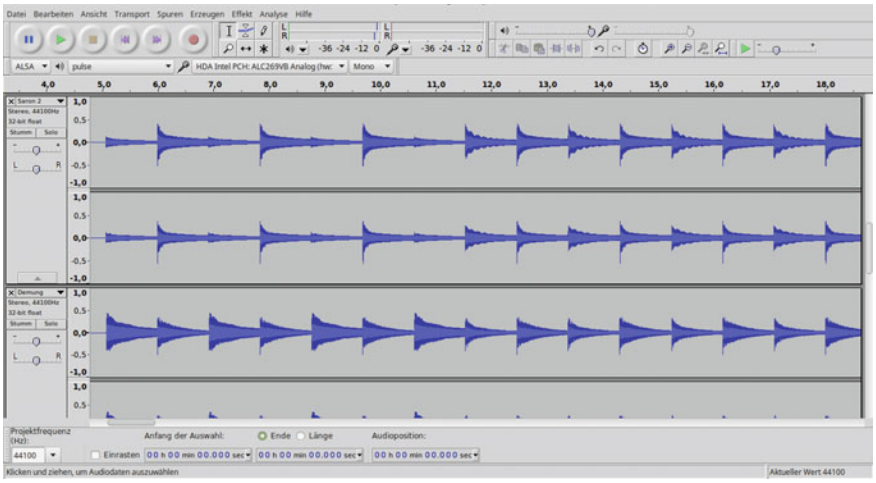


Fig. 2 Arranging the sound samples of single musical instruments with a DAW. The samples of the musical instruments were arranged regarding their role within the musical piece *lancaran manyar séwu. saron 1, saron 2, demung* realize the skeleton melody *balungan* shown in the notation below. *Peking* plays every *balungan* note twice. *Bonang barung* plays in octave style *gembyang* and anticipates the *balungan* notes. *ketuk, kenong* and *gong suwukan, gong ageng* realize the colotomic structure of the musical form *lancaran*

to their function within a traditional *gamelan* composition using a DAW (Fig. 2). For this step the well known gamelan piece *lancaran manyar séwu* was chosen. The sequenced version was based on the tones of the musical instruments, had a duration of 38 seconds and encompassed one gong cycle of the above chosen piece, as well as the skeleton melody *balungan* of the piece played once.

The Samples of following Instruments where included: *Saron 1, Saron 2, Demung, Peking, Bonang Barung, Kenong, Ketuk, Gong Suwukan, Gong Ageng*

2.1.2 Pitch Shifting

The single sound samples of each single instrument was pitch shifted several times to meet the pitches of the different tunings presented below. Building the musical piece from them in the DAW as described above different tuned versions of the same piece were built. For this process we used different pitch shifting algorithms that all shift the fundamental frequency including all spectral content to a desired frequency. Following tunings were built and used as stimuli in the following listening test:

2.2 Measurements of Auditory Roughness

For the measurements of auditory roughness of the stimuli, the Helmholtz/Bader algorithm was used, that already was part another study to measure the auditory roughness of synthesized organ corals [17].

Helmholtz proposed a definition of roughness using the amplitude modulations of all possible pairs of frequencies present in the sound. He defines a maximum roughness value with a distance of 33 Hz between two frequency values of $| f_1 - f_2 |$. Roughness is zero for $f_1 = f_2$ and is declining for $| f_1 - f_2 | > 33$ Hz to zero again asymptotically. He additionally weights the roughness with the amplitudes of the two frequencies. Also the roughness is independent from the absolute frequencies considered, and so $| f_1 - f_2 |$ results in the same roughness no matter if i.e. $f_1 =$ 20 Hz or if it is $f_1 = 20\,000$ Hz.

It is be interesting to implement this historical idea of Helmholtz as it is most simple and clear, and it was already successfully being tested with a musical example of 23 chords [17]. The chords were played once in a tempered and once in a pure tone tuning. As a result, the pure tone set was calculated as less rough than the tempered one.

To implement the algorithm it was needed to decide to measure the amplitude modulation in the sound directly or to use the frequencies. As only small, but still perceivable differences in roughness need to be detected, the frequency version was implemented, with a frequency precision of two digits behind the comma. This assures for static sounds present in this study a high precision of the amplitude modulation speed. So first all m partials present in the sound louder than a threshold of -46 dB in relation to the loudest partial with 0 dB are found and stored with their precise frequency and amplitude values. Then a matrix of m $\times$ m roughness values are calculated, where n = $m^2/2 - m$ are non redundant and used for the roughness estimation.

To calculate this Helmholtz roughness, the mathematical formula Helmholtz has used needed to be guessed. As most parameters of it mentioned above are known, this was not a difficult task. Then the formula for a single frequency pair n consisting of f_1 and f_2 $R_n = A_1 A_2 \frac{|df_n|}{f_r} e^{-|df_n|/f_r}$ is used where A_1 and A_2 are the amplitudes of the two frequencies f_1 and f_2 respectively, $df = f_1 - f_2$, and f_r = 33 Hz is the maximum roughness frequency difference which is fixed. e^{-1} is a scaling parameter to fix the maximum roughness to R = 1 for $A_1 = A_2 = 1 = 0$ dB, the maximum amplitude. As mentioned above, all amplitudes are set to $0 \leq A \leq 1$. So each frequency pair can contribute a maximum of R = 1 to the overall roughness which is calculated like $\mathbf{R} = \sum_{n=1}^{N} R_n \backslash$.

Figure 3 shows the roughness dependency on the frequency difference d_n.

As the stimuli were about 15 seconds long, a roughness value was calculated for each second, integrating all frequencies within this second and using it in the above formula to calculate the rouhgness for this second. As the lancaran pieces are quite slow in tempo, the one second integration is a reasonable value. To end at one roughness value for each piece, the mean of all roughness values was taken.

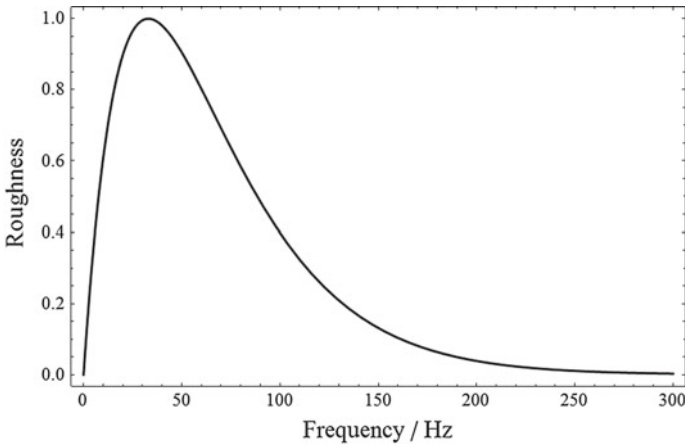


Fig. 3 Roughness depending upon frequency difference df for a single frequency pair f_1 and f_2 with $df = f_1 - f_2$. Maximum roughness appear at $|df| = 33$ Hz, zero roughness is at $df = 0$

2.3 Survey

The Stimuli were used as part of an online survey that was distributed to mailing lists related to the topic and directly to international *gamelan* groups. In order to test the subjective perception of the stimuli, the same dichotomous questions were asked for every tuning. In relation to the perception of dissonance and consonance, we chose the terms *out of tune* and *comfortable*. Following formulations were used:

Question A: Does the *gamelan* sound out of tune?

Question B: Can the overall sound be described as comfortable?

The questions were to be answered with the binary options yes/no. Subjects where asked to listen to the stimuli via a digital audio player that was integrated into the survey. It was possible to repetitively listen to the stimuli. Subjects where able to give optional additional information of their choice at the end of every page.

2.4 Correlation Between Auditory Roughness and Perception

In Order to relate the results of the auditory roughness measurements with the results of the survey, for each stimulus the amount of the binary variable yes for each question was taken as one variable, while the mean auditory roughness was taken as the other variable. To test a possible significant relation between both variables, the Pearson correlation coefficient was used.

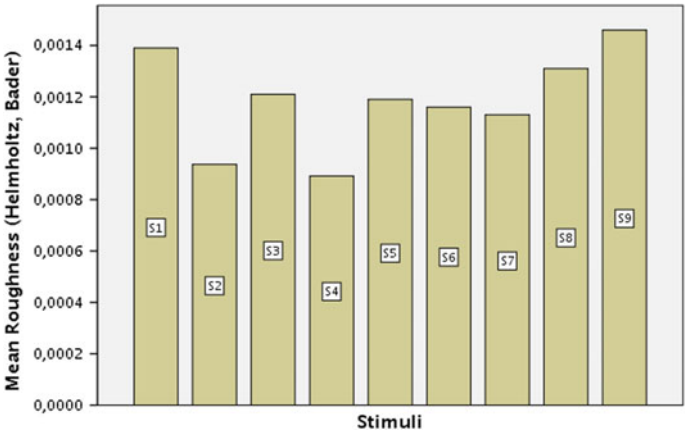


Fig. 4 Mean roughness of the nine versions of the original as well as the detuned *gamelan* piece built in a DAW from sampled *gamelan* instruments as calculated by the Helmholtz/Bader algorithm. The original version S1 is the roughest version next to the randomly detuned version S9. The versions tuned to theoretical perfect intervals S2 (equal tempered) and S4 (5-tone equidistant) are least rough

3 Results

3.1 Auditory Roughness of Stimuli

Figure *4* shows the mean roughness of the nine versions of the original as well as the detuned *gamelan* piece built in a DAW from sampled *gamelan* instruments as calculated by the Helmholtz/Bader algorithm. The original version S1 is the roughest version next to the randomly detuned version S9. The versions tuned to theoretical 'perfect' intervals S2 (equal tempered) and S4 (5-tone equidistant) are least rough.

3.2 Survey Results

3.2.1 Sample

The subject sample consisted of 27 participants (20 males, 7 females) 24 of whom indicated to already have participated in a *gamelan* group for a period of at least a year. 15 subjects indicated that they were never enrolled in or were at the time of research pursuing a music related major. The distribution of the nationality of the subjects was as follows: (USA: 11, Indonesia: 4, Great Britain: 3, Germany: 3, Japan: 1, New Zealand: 1, Netherlands: 1).

3.2.2 Evaluation of Stimuli

Figures 5 and 6 display the results of question A and B for all stimuli. The results of question A show that the original *pelog* tuning S1 was described by the least number of participants as out of tune, followed by the equal tempered tuning and the equidistant tuning.

The pseudo *hijaz* tuning S8 and detuned *pelog* S9 where described as out of tune by a largest number of subjects. It is interesting to remark that the equidistant tuning S4 and the *slendro* tuning S3, even though having frequencies very close to each others were rated quite differently by the participants. *Slendro* was described by a far larger number of participants as out of tune.

The results of question B shows similarities to the results of question A. However, the results of the same binary variables of both question are contrary. Therefore the results of question B are presented regarding the number of participants that answered the question B positive. The original tuning S1 was rated by the least number of participants as comfortable, followed by the equal tempered tuning and the equidistant tuning. The *pseudo hijaz* tuning S8 and the detuned *pelog* S9 was rated by the largest number of participant as not comfortable.

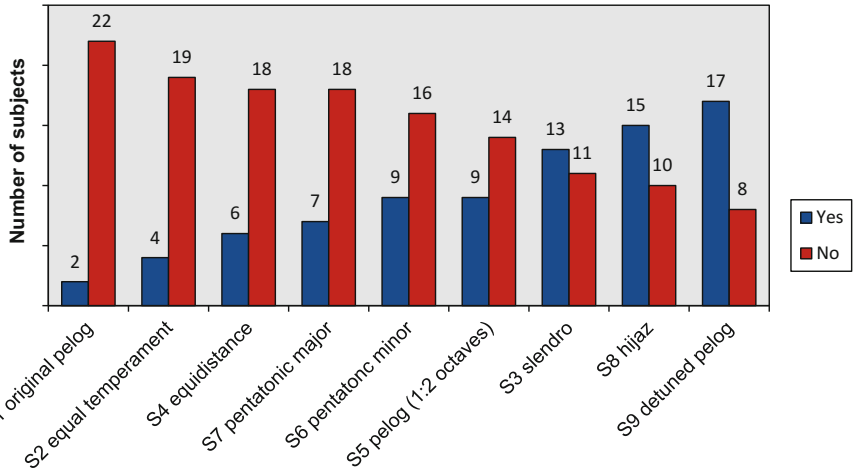


Fig. 5 Results of question A "Does the *gamelan* sound out of tune?". The stimuli are sorted by the amount of subjects that answered the question with no. The original tuning was described by the smallest number of participants as out of tune, followed by the equal temperament tuning. The detuned version S9 is described by the highest number of subject as out of tune

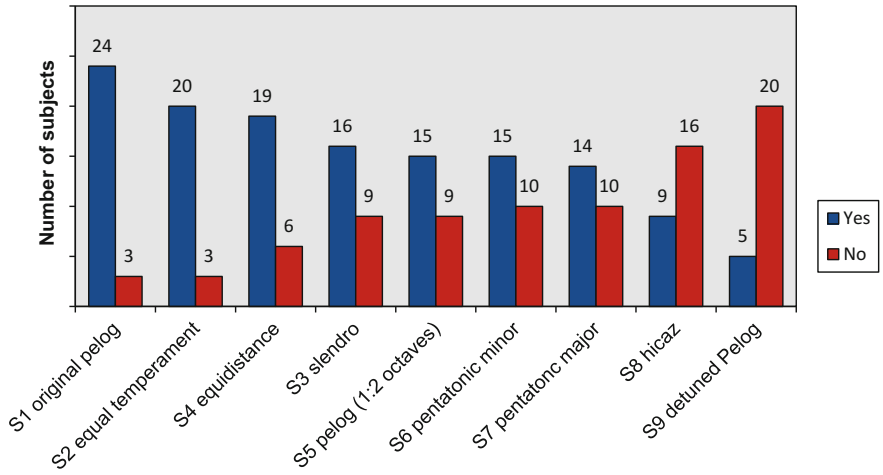


Fig. 6 Results of question B Can the overall sound be described as comfortable? The stimuli are sorted by the amount of subjects that answered the question positive. The original tuning is described by the largest number of participants as comfortable, followed by the equal temperament tuning S2. The ranking shows similarities to the results of question A

3.3 Correlation Between Results

Figure 7 displays the relationship between the two variables, amount of people that answered Question 1 and 2 positively and the mean auditory roughness for all nine stimuli.

The graph indicates that stimuli 2–9 follow a trend, but S1 stands on its own. Indeed, the Pearson correlation coefficient between calculated and perceived roughness resulted in no significant correlation.

Therefore it was decided to treat S1 as an outlier and exclude it from the calculations. Then the Pearson correlation between the two variables *out of tune* and *mean roughness* for stimuli 2–9 resulted in a highly significant positive correlation ($r = 0.927$, $p = 0.001$, $n = 8$). For the variable *comfortable sound* and *mean roughness,* a highly significant negative correlation ($r = -0.841$ $p = 0.009$, $n = 8$) was found. In summary, for stimuli 2–9 the amount of subjects that judged the tuning as *out of tune* increases with a higher mean calculated auditory roughness of the stimuli. The number of people that judged the sound of stimuli as *comfortable* decreases with a higher mean calculated auditory roughness value of stimuli.

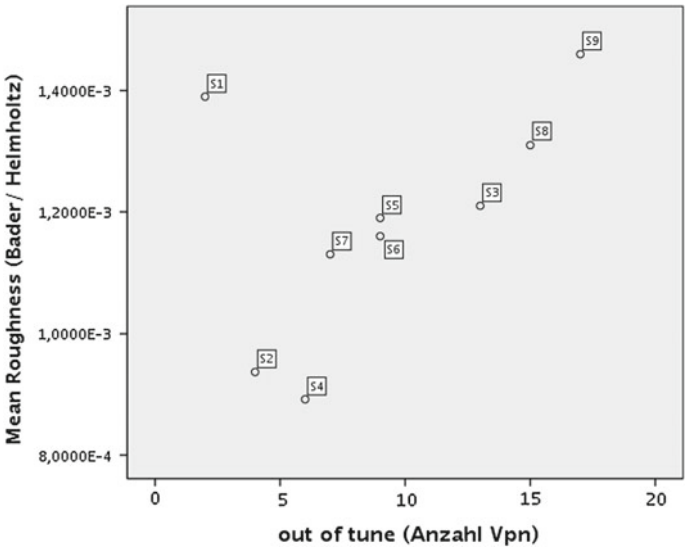


Fig. 7 The scater plot is relating the amount of subjects that described the stimuli as out of tune within the survey with the measured auditory roughness mean

4 Discussion

If the traditional tunings of *gamelan* ensembles were the result of the combination of the spectrum of the instruments with the minimal points of auditory roughness, we would expect the traditionally tuned stimuli to have a lower auditory roughness value. In contrast to that assumption, the original *pelog* tuning has a high roughness value compared to the other stimuli. The equal tempered tuning, even though having frequencies close to the original tuning, has a lower auditory value. The same phenomena, can be observed between other stimuli. Even though the frequencies of the *slendro* and equidistant tuning are quite close, *slendro*, despite being a traditional *gamelan* tuning, has a higher auditory roughness value than the equidistant tuning. The survey results suggest, that the original tuning, despite having a high roughness mean, was by the majority of subjects perceived as *in tune* and *comfortable* and that the equal tempered tuning was perceived as being quite similar even though it has a much lower measured auditory roughness value. Those findings do not support the view that traditional *gamelan* tunings can be explained with the minimal points of auditory roughness in combination with the spectrum of the instruments. By relating the results of the survey to the measured means of auditory roughness, it could be shown however, that in the cases of stimuli 2–9 significant correlations exist, that might be interpreted as a positive relationship between the perception of dissonance and auditory roughness. The results are not significant anymore if the original tuning S1 is included.

The relatively high roughness value of the original *pelog* tuning might be best explained by the small tuning varieties between instruments, that naturally occur within *gamelan* ensembles due to fact that the instruments are traditionally tuned by ear, while the tuning changes due to different conditions [7, 16]. However, the survey results suggest that S1 was perceived relatively consonant. Those findings show similarities to a study by Schneider et al. [17] that found that subjects preferred the equally tempered version of a synthesized organ coral over the justly tuned version, even though it had a higher auditory roughness. The authors conclude that the listening habit of the subject might influence the perception of auditory roughness. Since our study was mainly done with subjects already exposed to *gamelan* tunings, it seems possible that a certain amount of auditory roughness is part of the *gamelan*natural excepted sound.

To further emphasize this point, we want to include a free comments that was given by one subject of our study about the difference between the original tuning (A) and the equal tempered tuning (B):

> The instruments of B sound more in tune with each other… is that right? Yet the tuning of A is more Javanese and hence nicer to listen to. And maybe even having a few tuning variances between instruments actually makes it nicer to listen to?

However, we cannot conclude that listeners playing *gamelan* are used to the *gamelan* tuning and therefore prefer it simply because they know and like it. Explaining the results by only a habit could only work when all *gamelan* ensembles the subjects play and know and the original tuning S1 presented in the study were the same. Still each *gamelan* has its own tuning. Therefore the subjects were not used to S1 and therefore this line of reasoning cannot be the cause of the preference of S1. Still as indicated in the quote above, the subject was able to identify S1 as a typical *gamelan* tuning. Taking the large tuning deviations of different *gamelans* and the rejection of S2–S9 into consideration this identification is astonishing.

One hypothesis to test in the future could be that the instrument tuner tries to fit the tuning of the gamelan set according to the respective overtone spectra of this special set in order to reach a ‘typical’ *gamelan* sound. Then such a typical *gamelan* sound would be a combination of both, a the overtone spectra and the tuning. Obviously there need to be a certain amount of roughness to make the sound interesting. Still it might be that roughness need to be taken not as a single value but differentiated in terms of a roughness texture. Just as there is not only one distortion of electrical guitars. Such distortion can also be measured with a single value, still Rock history knowns an endless amount of distortions styles and textures.

5 Conclusion

This following study examined the influence of auditory roughness on the perception of different *gamelan* tunings. This was done by measuring the auditory roughness as

well as testing the perception of different *gamelan* tunings that where created through pitch shifting and sequencing of recorded sound samples. The results indicate that the role of auditory roughness in the case of *gamelan* tunings seems to be more complex than just inducing dissonance. For further research it would be beneficial to have a larger sample size, as well as subject groups with different levels of experience in the field. Additional measurements of auditory roughness as well as spectral content of existing *gamelan* ensembles in different tunings could provide a further insight into the topic.

Acknowledgements The instruments whose tones were recorded for the stimuli of this study are part of the Gamelan Margi Bodoyo that belongs to the Indonesian Consulate of Hamburg (KJRI). We thank the Consulate for their support.

References

1. Sutton, R. Anderson. (1985). *Musical Pluralism in Java: Three Local Traditions.* Ethnomusicology, Vol. 29 (1), 56–85
2. Pickvance, Richard. (2005). *A Gamelan Manual: A Player's Guide to the Central Javanese Gamelan,* London: Jaman Mas Books.
3. Hood, Mantle. (1980). The Evolution of the Javanese Gamelan: Book 1: Music of the Roaring Sea. Heinrichshofen.
4. Kieven, Lydia (2006). Sound and movement in stone - music and dance in ancient Javanese art. In: Andreas Lüderwaldt (eds.), *Contemporary Gamelan Music: 3. Internationales Gamelan Musik Festival Bremen 2006*. Jahrbuch XIV Überseemuseum Bremen. (pp. 9–22). Bremen.
5. Sumarsam. (1981). *The Musical Practice of the Gamelan Sekaten*. Asian Music, Vol. 12 (2), 54–73
6. Brinner, Benjamin. (1993). *A Musical Time Capsule from Java. Journal of the American Musicological Society*, 46, (2), 221–260.
7. Vetter, Roger. (1989). *A Retrospect on a Century of Gamelan Tone Measurements.* Ethnomusicology, 33, (2), 217–227.
8. Hornbostel, E. M. (1927). Musikalische Tonsysteme. In H. Geiger & K. Scheel (Eds.), *Handbuch der Physik* (8th Edn., pp 425–449). Berlin: Springer.
9. Sethares, William A. (2004). *Tuning, Timbre, Spectrum, Scale*. Berlin: Springer.
10. Fletcher, Harvey, 'Auditory Patterns', Rev. Mod. Phys., 12, (Jan, 1940), 47–65.
11. Helmholtz, Herman von (1863), *Die Lehre von den Tonempfindungen als physiologische Grundlage der für die Theorie der Musik*, Braunschweig: Vieweg (3. Aufl. 1870, 5 Uaf. 1896, 6 Aufl. 1913)
12. Vasilakis, Pantelis N. (2005). Auditory Roughness as a Means of Musical Expression. In Kendall, Roger A. & Savage, W.H (Eds.) Selected Reports in Ethnomusicology, 12: 119–144 (Special Issue: Perspectives in Systematic Musicology)
13. Ramaer, Huib. (2004). Sinta Wullur and the Diatonic Gamelan. Balungan, 9–10, S. 30–33.
14. Suppangah, Rahayu. (2003). Campur Sari, A Reflection. Asian Music, Vol. 34 (2), 1–20.
15. Alves, Bill. (1997). "Pleng: Composing for a Justly Tuned Gender Barung," Presented at the Fourth International Symposium and Festival of Intercultural Music, London, April 1996, Journal of the Just Intonation Network 1, 4–11.
16. Surjodiningrat, Wasisto, et al. (1972). *Tone Measurements of Outstanding Javanese Gamelans in Jogakarta and Surakarta*. Second ed., Jogjakarta: Gadjah Mada University Press.
17. Schneider, Albrecht., Ruschkowski, Arne., & Bader, Rolf. (2009). *Klangliche Rauhigkeit, ihre Wahrnehmung und Messung*., In Bader, Rolf. (Eds.) Musikalische Akustik, Neurokognition und Musikpsychologie. Hamburger Jahrbuch für Musikwissenschaft, 25, (pp. 103–149) Frankfurt am Main: Peter Land

Part III
Computation, Networking and Platforms

Content-Based Music Retrieval and Visualization System for Ethnomusicological Music Archives

Michael Blaß and Rolf Bader

Abstract In this chapter we propose a content-based exploration and visualization system for ethnomusicological archives that allows for data access by rhythm similarity. The system extracts an onsets-synchronous timbre feature of each audio file of a given collection. From the resulting time series, Hidden Markov Model are trained. The transition probability matrices of the models are considered a rhythm fingerprint that represents the musics rhythmic structure in terms of timbre. The self-organizing map algorithm is utilized to project the high-dimensional fingerprints onto a two-dimensional map. This technique preserves the topology of the high-dimensional feature space, which results in similar map positions for similar rhythms. A clustering by rhythm similarity is thus achieved. The system, therefore, supports musicologist studies in several ways: the rhythm fingerprinting does neither imply a certain theory of music nor introduce cultural bias. Hence, different musics can be compared meaningfully regardless of their origin. Retrieval by similarity allows for an explorative approach to the music collections, which can support researchers in finding new hypothesis and utilizing music archives with few or without meta data. The system is currently prototyped in the Ethnographic Sound Recordings Archive of the University of Hamburg as a part of the COMSAR project.

1 Introduction

Recently, a lot of effort has been put into digitizing ethnographic audio collections. This endeavor features two aspects: *preservation* and *utilization*. The first aspect addresses the conservation of vintage and rare audio material, which may be only available on decaying media such as wax cylinders, shellac or acetate discs, as well as the safeguarding of cultural heritage via ethnomusicological field work. The African

M. Blaß (✉) · R. Bader
Hamburg, Germany
e-mail: michael.blass@uni-hamburg.de

R. Bader
e-mail: rolf.bader@uni-hamburg.de

R. Bader (ed.), *Computational Phonogram Archiving*, Current Research in Systematic Musicology 5, https://doi.org/10.1007/978-3-030-02695-0_7

Music Archives[1] of the Johannes Gutenberg University Mainz features about 10,000 records of African popular music. Currently, the collection's meta data is being gathered into an online catalogue. This effort allows for text based search and retrieval, without access to the audio data. The Dutch Song Database[2] provides access to 140,000 scores of dutch songs with a constantly growing number of attached song texts and audio recordings [1]. Search and retrieval is purely text based. Additionally, the archive provides access recorded sounds. The Telemeta project[3] provides access to more that 21,000 audio recordings of the Center for Research in Ethnomusicology[4] (CREM) [2]. Telemeta combines text as well as geographic search and retrieval methods with an audio player. The player allows for streaming audio data as well as displaying visual information such as waveform, spectrogram and other sound features.

The second aspect addresses the utilization of ethnographic audio collections once they are digitized. Even though this aspect did not receive much attention, we like to emphasise that the true value of audio collections for ethnomusicology emerges from their usage. However, collections can only be used if the owners provide access in manner that is suitable for ethnomusicological research. For example, the commonality of the above mentioned projects is the focus on individual recordings. It is possible to query the data base for some keywords, display and possibly listen to the results. It is not possible to compare recordings or their meta data side by side. However, this is an essential feature for ethnographic research. The Digital Music Lab[5] provides smart way to compare pieces automatically on large scale. The project enables users to compare aggregates of extracted audio features of sections of pieces, whole pieces and even whole data bases [3]. This is an interesting approach, which supports research that focuses on the global structure of music archives. The user has, however, only access to the features. Access to audio files and hence listening, is not granted in order to circumvent copyright infringement. However, listening is essential to ethnomusicological research. Thus, tools have to be provided that allow for meta data based search on and audio feature based workflows.

In this chapter we introduce an automatic organization and visualization system for ethnographic music archives, which is currently being prototyped as part of the *Computational Music and Sound Archiving Project* (COMSAR). The goal of the project is to integrate classical, ethnomusicological workflows and data-driven approaches in a web-based front end. The former is designed as a rich and intuitive meta data structure specific to ethnomusicological records. The second part provides of a content-based visualization system based on the Self-Organizing map. Figure 1 depicts the user interface of the COMSAR system.

The underlying feature extraction system utilizes Hidden-Markov models to compute a song-level rhythm feature based on timbre. This purely data-driven approach

[1] https://www.ama.ifeas.uni-mainz.de/.

[2] http://www.liederenbank.nl/.

[3] http://www.telemeta.org.

[4] http://crem-cnrs.fr/.

[5] http://dml.city.ac.uk/.

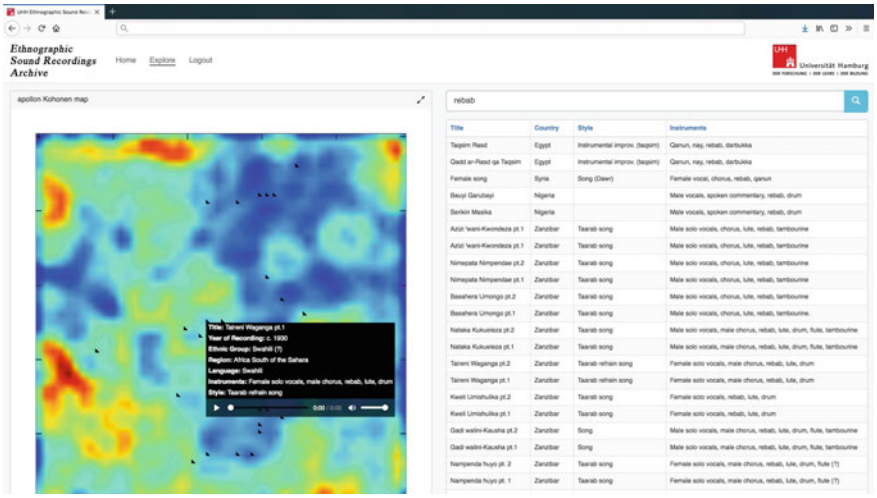


Fig. 1 User interface of the COMSAR system as implemented in the Ethnographic Sound Recordings Archive of the University of Hamburg

is not biased regarding any music theory and is hence suitable for ethnomusicological research. The self-organizing map depicts similarities inherent in the archive structure as distances on a two-dimensional map. This allows for tow different workflows: using the text input field the user can browse an archive by meta data. All available meta data for each result as well as an audio player is provided on-click. Additionally all results to a query are marked on the self-organizing map. Hence, similarity can easily be assessed by distance on the map. By default the map represents markers for all records in the archive. The map background provides cues to the cluster structure inherent in the archive. Hence, the user can explore the archive based on visualization. The markers on the map provide a subset of the available meta data and a simple audio player, such that predictions of the algorithm can be evaluated. This functionality is extremely useful for the exploration of old archives, which may come with few or without meta data. Furthermore, it is possible to compare newly recorded sounds with the archive by simply uploading the file. The new sound will then be marked on the map such that the user is able easily retrieve similar sounds. Such a search by similarity may support generating new hypotheses fast during filed trips.

The remaining chapter is organized as follows: Sect. 2 provides an overview of automatic organization of music archives and a discussion of specifics of ethnographic music archives. Section 3 features a detailed introduction to the COMSAR rhythm extraction system. The following Sect. 4 focuses on the organization of features by the self-organizing map. Section 6 gives a complete description of the

COMSAR system as it is currently implemented in the Ethnographic Sound Recordings Archive of the University of Hamburg. Subsequently, an example application is provided in Sect. 7, which is follow by some concluding remarks in Sect. 8.

2 Ethnomusicological Sound Archives

2.1 Automatic Organization of Music Archives

During the last two decades, numerous methods for the automatic content-based organization and classification of music archives have been proposed. In [4] a methodological framework as well as one of the first toolboxes for content-based music retrieval is provided. In [5–7] content-based retrieval systems for melodic phrases in MIDI the format are proposed. Furthermore, it was demonstrated that self-organizing maps can be applied to automatically organize music archives by genre. The authors of [8, 9] trained a feature map on a spectral representation of segments of five seconds length, which were extracted from the archived audio files. Since the spectral content of a musical piece varies over time, the segments that belong to one song are scattered over the map. For each audio file, a feature vector was constructed from the positions of its segments on the feature map. From the position vectors, a second map was trained, on which a clustering by genre was obtained. This work was improved by extracting the fluctuation strength, from a 30 s segment from the middle of each song [10]. The underlying assumption is to capture the most significant features, which are thought to reside in the temporal center of the songs. This approach is questionable, given a data set of western popular music. In the context of ethnographic music archives, it is clearly not applicable. The distribution of interesting features depends on the music culture and the circumstances of recording. Hence, they could also be found at the beginning, at the end, or even both. The self-organizing map was also utilized to explore the melodic contours of folk song cultures from Europe, Asia, and North America [11, 12]. The resulting clustering revealed contours that are typical for each culture. Furthermore, the arrangement of the clusters allowed conclusions about the musical cultures.

2.2 Problems Specific to Ethnomusicological Archives

2.2.1 Ground Truth Annotations

Music information retrieval systems for ethnographic archives are often evaluated by classification task, where the system predicts class labels for each item in the archive. From the systems' predictions and a given ground truth, classification scores are computed. In [13] a study is provided in which the authors test three classifiers on

a world music subset of the Smithsonian Folkways Recordings. The subset includes 50 songs form each of 31 featured countries. Country names were used as ground truth assuming that music from the same country shares a common set of features. They achieved a maximal classification accuracy of 40.6% using Linear Discriminant Analysis to obtain the feature space on which classification was performed using k-Nearest Neighbors, with $k = 3$. This result indicates that automatic classification by country does not seem to be an adequate approach to explore ethnographic music archives. Indeed the author's assumption is not valid. Consider, for example, the pieces *Helv el mabassem* [14] and *Taxim Rast* [15] from ESRA. They both are recorded in Egypt but are neither similar in instrumentation nor in style. Hence, the annotations used as ground truth are problematic. They should at least reflect ethnic groups to be useful for ethnomusicologist. This is especially desirable for regions where ethnicities live across national borders.

2.2.2 Availability of Meta Data

Contemporary music collections, as well as fieldwork, commonly features a multitude of meta data. Vintage record collections lack this abundance of information, as it is the case with ESRA (see Table 1). The availability of meta data is, however, crucial for any classification task since a ground truth relies on a complete set of annotations.

2.2.3 Unbalanced Data

Another meta data related problem arises from the ratio of samples to the number of classes. To train a classifier it is necessary to have an adequate number of examples per class. This is usually not the case with archives of field recordings. There are commonly only a few samples for a large number of classes. It may also be the case that one features considerably more samples as the remaining classes. This also the Wilhelm Heinitz Collection of ESRA: there more then forty recordings from Algeria, twenty three from the Ivory Coast, and only two from Syria.

2.3 Summary

Ethnographic archives typically contain fieldwork recordings. It is of great interest to ethnomusicologist to organize these by music similarity. Tag-based approaches may isolate certain features. However, there is no confidence that they indeed reflect music similarity. Hence, content-based approaches are to be preferred. Many of these approaches focus on classifying western popular music by genre. Consequently, their performance depends on culture. This is not desirable in ethnographic archives. The

system has to deal with, for example, drum language of West Africa and ancient Japanese chant.

A further problem is that features are only taken from a distinct segment of each record. Hence a feature extraction method is desirable, that aggregates local low-level features to a global high-level. To be comparable in clustering algorithms, the feature size has to be invariant regarding the duration of the records.

3 Timbre-Based Rhythm Theory

In this section we introduce our approach of modeling the rhythm of a given piece of music in terms of its constituting timbres, which was firstly introduced in [16, 17]. First we present an outline of the theory. Afterwards, a short introduction to Hidden Markov models, the computational core of the theory, is given. This is followed by a detailed description of the modeling process.

3.1 Theory

Definitions of musical rhythm, which serves as basis of rhythm extraction models, commonly incorporate only regular partitions of time as atomic elements [18]. A particular rhythmic pattern is then given as a succession of regular time intervals, with their limits representing musical events, that is, note onsets or rests. The perception of a rhythmic pattern as a whole, the *groove*, can change if the musical events change context. Compare, for example, the pattern (a) of the score in Fig. 2 with pattern (b). The only difference is that bass and snare drum swapped their position after the rest. This rather slight change in the score has a significant effect on perception. The upper example conveys a feeling of progressive movement, whereas the second sounds stumbling. This is due to the rather unconventional placement of the snare drum on beat 3+. Yet rhythm theories would not account this as a change in rhythm, compared to the upper example. The third pattern in Fig. 2, compared to both examples (a) and (b), however, would account for a change in rhythm and perception. Theories would argue, that this is because the inter-onset intervals changed (rest on 2+ an 4+) and one note is missing.

We argue that this phenomenon can be handled more comprehensively by explaining it in terms of timbre. Each note onset is not only treated as an event in time. A second dimension is attributed to it, which describes the sound quality of this event, that is timbre. The comparison of patterns (a) and (c) results in a rhythmic difference because of two changes. (1) The edits: on beat 3+ a timbre with high energy in the

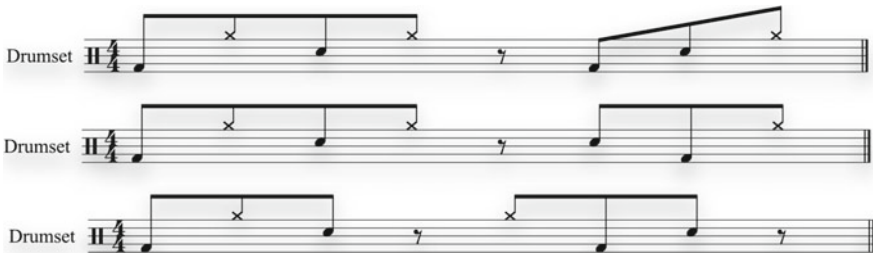


Fig. 2 Drum pattern examples. From top to bottom: **a** a baseline pattern. **b** Minor modifications alter perception but do not count as rhythmic change. **c** Modifications incorporating rests do count as change is rhythm

lower end of the spectrum was substituted with a timbre that has more energy in the upper range. On beat 4 the inverse operation was applied. (2) Dependance of timbre: organs of perception work sequentially. Physical stimuli that arrive in a certain order are aggregated to a sensation. In auditory perception this order is crucial since an expectation about the next event is established given the present and past events. That is, successive rhythmic events are not independent of each other. The expectation about *when* future onsets are going to happen has already been examined [19]. We propose a theory that models the expectation of *what* is going to happen next given the dependence between successive events.

We argue that the percept of a rhythm pattern as a whole is partly established by a distinct succession of polyphonic timbres [20]. 'Polyphonic' refers to the fact that auditory sensation may emerge from a composition of several sounds. Hence, at each onset there may be the timbre of one instrument or a mixture of timbres of multiple instruments. Using the Hidden Markov model it is possible to aggregate the transitions between the single timbres in a musical piece. Similar timbres, that differ because of the playing technique (snare stroke vs. rim shot) are thereby grouped together. The result can be interpreted as rhythmic fingerprint.

Such models have the useful property to be independent of wrap around. Given a constant, repeating groove of the timbres $G = (t_1, t_2, t_1, t_3)$ and a shifted version $G\prime = (t_3, t_1, t_2, t_1)$. Using the same hyper parameters for training, the resulting models would be exactly the same. Hence, one can expect that missing values alter the model slightly, whereas the overall structure remains the same (see Fig. 3). Therefore, the model is robust against recording errors, such as delayed start of recording.

The view on musical rhythm from the point of timbre bears several advantages over the traditional view, which only incorporates inter-onset intervals. It allows to explore polyrhythmic patterns in a comprehensive manner. As overlapping sounds are modeled as mixture of timbres, the emerging patterns of polyrhythms can be described as a succession of polyphonic timbres. It furthermore allows to model melisma and figurations that are not heard as single tones like, for example, in bagpipe music.

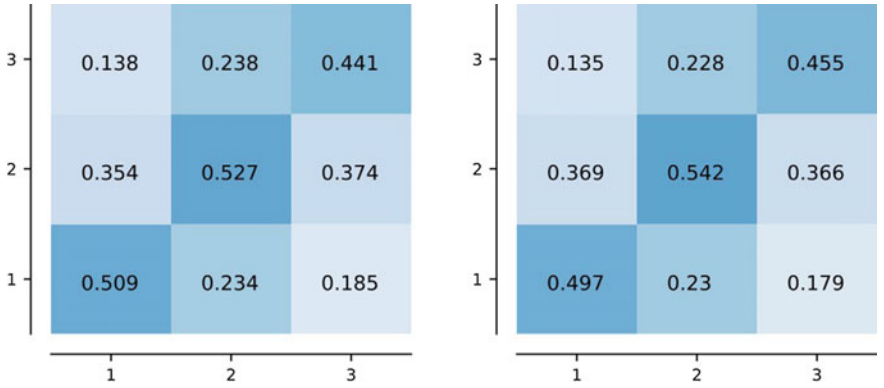


Fig. 3 Two models of ESRA sample *EAn025-A1* trained with the same hyper parameters. Left: trained on full time series (285 values). Right: first 20 samples removed

3.2 Introduction to Hidden Markov Models

Consider a sequence of observed symbols $O_T = \{o_1, o_2, \ldots, o_T\}$. In the Hidden Markov model (HMM) it is assumed that each value $o_t \in O_T$ is a sample drawn from a stochastic process $\{X_t : t \in \mathbb{N}\}$, which is distributed according to a mixture model with m component distributions. The transitions between the components are modeled as transitions of the state of a stochastic process $\{C_t : t \in \mathbb{N}\}$, which is considered to be a Markov chain (MC). Markov chains change their state according to a transition probability matrix

$$\mathbf{\Gamma} = \begin{pmatrix} \gamma_{11} & \cdots & \gamma_{1m} \\ \vdots & \ddots & \vdots \\ \gamma_{m1} & \cdots & \gamma_{mm} \end{pmatrix} , \tag{1}$$

where the elements of $\mathbf{\Gamma}$ represent the probabilities of changing to state j given the actual state is i. If $\mathbf{\Gamma}$ is constant over time, the Markov chain is said to be homogeneous. In this case the elements (i, j) of $\mathbf{\Gamma}$ are defined as the conditional probabilities

$$\gamma_{ij} = \Pr(C_{t+1} = j \mid C_t = i) \quad . \tag{2}$$

For the remainder of this chapter all HMMs are considered to contain a homogeneous Markov chain as unobservable process. Given Eq. 2 the transition to state C_{t+1} depends only on C_t. Hence it holds that

$$\Pr(C_{t+1} \mid \mathbf{C}^{(t)}) = \Pr(C_{t+1} \mid C_t) \quad , \tag{3}$$

where $\mathbf{C}^{(t)} = (C_t, C_{t-1}, \ldots, C_1)$ denotes the history of the process $\{C : t \in \mathbb{N}\}$ from the start until time step t. Equation 3 is known as the Markov property. Since the states of the MC each correspond to a component distribution of the mixture model, X_{t+1} only depends on C_t, too. That is, X_{t+1} does neither depend on previously observed symbols $\mathbf{X}^{(t)}$ nor on the history of states $\mathbf{C}^{(t-1)}$. The Hidden Markov model is hence characterized by

$$\Pr(X_{t+1} \mid \mathbf{X}^{(t)}, \mathbf{C}^{(t)}) = \Pr(X_{t+1} \mid C_t) \quad . \tag{4}$$

The model rejects the assumption of independence in favor of a serial dependence. Hence, the HMM is a particular kind of dependent mixture model, in which dependence appears only between consecutive samples of the unobservable Markov chain.

Under the HMM a sequence is analyzed by estimating the model parameters. The number of parameters depends on the number of states m, which is a hyperparameter that has to be specified in advance, and the number of shape parameters the component distributions take. This gives a total of $m^2 + m$ model parameters. These are usually estimated by maximizing the likelihood L_T of the HMM regarding a vector of parameters $\boldsymbol{\Theta}$ given a sequence of observations $\mathbf{X}_T$ with index $t \in 1, 2, \ldots, T$. The likelihood is generally given as

$$L_T(\Theta|\mathbf{X}) = \boldsymbol{\delta} \left(\prod_{t=1}^{T} \boldsymbol{\Gamma} \mathbf{P}(x_t) \right) \mathbf{1}_m^{\intercal} \quad , \tag{5}$$

where the row vector $\boldsymbol{\delta}$ represents the stationary distribution of the model. $\mathbf{P}(x_t)$ is the $m \times m$ diagonal matrix

$$\mathbf{P}(x_t) = \begin{pmatrix} p_1(x_t) & & \\ & \ddots & \\ & & p_m(x_t) \end{pmatrix} \tag{6}$$

of state dependent probability masses (or densities)

$$p_i(x) = \Pr(X_t = x \mid C_t = i) \tag{7}$$

and $\mathbf{1}_m$ is a row vector of m ones. For parameter estimation, Eq. 5 may be maximized numerically [21]. A more common approach is to treat the parameters as missing values and apply the Baum-Welch algorithm, which is a specialization of the well known EM algorithm.

Hidden Markov models are particularly useful to analyze dependent, discrete valued time series [21], which is why they are of particular interest in the analysis of feature vectors extracted from audio signals. Their stochastic nature additionally allows to compute distance metrics and hence retrieve no exact matches. HMMs have been successfully applied to a variety of problems in musicology and music information retrieval such as audio segmentation [22], rhythm analysis [23], genre

classification [24] intelligent music agents [25] and networked music performance [26]. A gentle introduction to HMM is given in [27] and a more comprehensive one with example applications can be found in [28].

3.3 Timbre-Based Rhythm Feature

In order to extract the timbre-based rhythm feature proposed in Sect. 3.1 it is necessary to perform a low-level feature extraction, which results in a description of timbre at musically relevant instants. We follow a similar approach as [29, 30], who modeled the "global timbre" of a musical piece as a whole. They extracted a timbre descriptor from consecutive short time windows of their pieces and modeled the resulting time series as a HMM. We are specifically interested in the "global rhythm" as a whole. Hence, we have to apply onset detection in the first place. This is necessary since we need to capture a most representative snapshot of each involved timbre. Because inter-onset intervals might be very short, the snapshots have to be taken right at the onsets. If the snapshots were taken from a sliding window of fixed length, we would accept quite a number of snapshots that represent horizontal timbre mixtures. Horizontal mixtures are the sum of parts of two successive sounds. In order to model the rhythm we need to aggregate only vertical mixtures, that is, instruments that play at the same time. The technical details of the onset detection and timbre feature extraction are given in Sect. 6.4.

3.3.1 Timbre Related Audio Features

In music information retrieval, the most widely used feature for timbre models are *Mel Frequency Cepstrum Coefficients* (MFCC). They constitute the de facto feature for automatic speech recognition since their introduction by [31]. MFCC extraction is a deconvolution technique, based on the independent source-filter model of speech. The goal is to separate the periodic source signal (glottis) from the filter (vocal tract), which shapes the spectrum, in order to allow for pitch independent speech recognition. The application to timbre models is straightforward: eliminating the periodic source signal (excitation) of an instrument from its filter characteristic (geometry, room), would immediate result in a description of the instruments timbre.

Recently the usage of MFCCs as a descriptor for musical timbre has been criticized, since there is no obvious relation between MFCCs and timbre perception. An adjective rating study concerning explicitly polyphonic timbres found that MFCCs did not at all correlate with one of the perceptual timbre dimensions they obtained [20]. The only exception is the 13th MFCC, which shows a weak correlation with a spatial dimension related to *fullness*.

Alternative features are motivated by psychological experiments. A common approach is to assess timbre perception by multidimensional scaling (MDS) of dissimilarity ratings of complex, synthesized or real instrument tones [32, 33] or of physical sound parameters [34]. Another approach is Factor Analysis (FA) of adjective ratings [35, 36] regarding sound stimuli. The motivation of both approaches is to condense the complexity of the participants' ratings to a low-dimensional *timbre space* such that the dimensions or factors correspond to the salient features the participants used to rate the stimuli. The solutions provide a means for visual inspection, qualitative and quantitative interpretation as well as clustering regarding the perceptual dimensions. Further methods involve semantic analysis of verbal sound descriptions (adjective ratings) and MDS of physical parameters.

In most cases studies obtained solutions of two to three dimensions. One of these was always related to a spatial feature within the stimuli, that is the distribution of spectral energy. The remaining dimensions have been interpreted as temporal dimensions correlating with the on- and offset patterns of spectral components and the presence or absence of hight frequency partials during the attack [37]. The weighted mean frequency of the spectral energy distribution, the spectral centroid (SC), was found to correlate strongly with the dissimilarity ratings along the perceptual dimension associated with the spectral energy distribution [38]. SC in turn correlates strongly with the perception of *brightness* [39, 40], making it a natural parameter for coarse discrimination of timbres. The spectral centroid of a spectrum is defined as

$$SC = \frac{\sum_i f_i A_i}{\sum_i A_i} \quad , \tag{8}$$

where f_i is the frequency of the ith bin and A_i is the corresponding magnitude.

We decided to choose the spectral centroid as one-dimensional model for polyphonic timbre for several reasons: (1) Literature suggests that spectral centroid is one of the most prominent features for dissimilarity ratings. (2) [41] addressed the different approaches to timbre models in music psychology and music information retrieval. In their extensive review they found that spectral centroid performs comparable to MFCCs in genre classification tasks. (3) Spectral centroid is a inexpensive to calculate scalar measure of the spectral energy distribution. MFCCs are vectorial quantities that also represent the spectral energy distribution, but in more detail. However, they are computationally expensive compared to SC. (4) Since MFCCs are vectorial, one has to decide how many MFCCs should processed. This is a hyper parameter that has to be set. However, we want to avoid manually set model parameters as much as possible.

3.3.2 Component Distributions

The choice of the component distributions depends on the problem under investigation. All probability distributions are possible candidates; it is even possible to choose instances from different distribution families for each state. Most studies in

MIR choose a mixture of possibly multivariate Normal distribution. Hence, each state is specified by a vector of means $\boldsymbol{\mu}_k$ and a covariance matrix $\boldsymbol{\Sigma}_k$ for each of K mixture components, such that

$$X_i \sim \sum_{k=1}^{K} \phi_k \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \quad , \tag{9}$$

with ϕ_k being a mixture parameter and $\sum_{k=0}^{K} \phi_k = 1$. This study is only concerned with the modeling of a scalar feature. Hence, the complexity of the component parameterization can be reduced to $X_i \sim \mathcal{N}(\mu, \Sigma)$. The spectral centroid is a frequency and hence represents the number of periods per second. This is a count. The standard model for count data in statistics is the Poisson distribution given by the probability function:

$$p_\lambda(x) = \frac{\lambda^x}{x!} \exp^{-\lambda} \tag{10}$$

The Poisson is a discrete probability distribution and shaped by only one parameter λ, which represents the mean. The Poisson has the additional property that its variance equals its mean. If the mean increase, the distribution becomes broader. We can exploit this effect. Filter banks, as with the MFCCs, have center frequencies, which are equidistant on the log scale to accommodate human pitch perception. The filter bandwidth is narrow for low frequencies since small frequency deviations in the lower band readily lead to a change in perception ($f_{B4} - f_{A4} \approx 53$ Hz). As the frequency increases, deviations also have to increase in order to alter perception ($f_{B5} - f_{A5} \approx 99$ Hz). The same applies for the Poisson distribution. For low λ, rather small deviations of x from the mean lead to low probabilities of x. The lower $\Pr(X_i = x)$ the lower is the probability that x belongs to state X_i. The Poisson is, therefore, a natural choice for our purpose.

3.3.3 HMM Training

Model training refers to the adaption of model parameters given a data set, such that the likelihood of the model (Eq. 5) is minimized. The model takes one hyper parameter, the number of component distributions m. It is possible to estimate m using Bayesian inference; however, this approach is rather elaborate and requires tremendous computational efforts. Furthermore, it is not necessary to estimate m. In nearly all cases there is enough prior knowledge to guess suitable vale. On the other hand, the number of component distributions roughly reflects the number of polyphonic timbres in the analyzed sound. Hence, it has a strong impact on the interpretability of the result.

In a HMM 20 parameters have to be estimated for a model with $m = 4$. It would be absurd to train such a model on a data set with $m \approx T$. For $m = 1$ the HMM simplifies to an m-state mixture model.

4 Data Visualization by Self-organizing Maps

The Self-Organizing Map is a well-known tool for dimensionality reduction, clustering and visualization. It has been applied to a variety of problems such as robotics, speech recognition, ecological research [42] finance [43], and automatic organization of text documents. Within the musical context, the SOM has been applied to folk music analysis [11], tonality perception models [44], music recommender systems [45], music control systems [46] and browsing of music archives [47]. In this section the SOM is briefly introduced as well as visualization techniques for SOMs.

4.1 Structure of the SOM

The SOM can be seen as special case of feed-forward neural networks. It consists of an input layer and a subsequent map layer. The input layer is determined by a set $\mathcal{M}$ of N input neurons (or units). The map layer has a certain topology (one-dimensional, rectangular, toroid, spherical). In the most common case the map units are placed on a rectangular and regular grid of dimensions d_x, d_y, with $N = d_x d_y$. On the map layer, units are laterally connected to neighborhood. Both layers are fully connected, that is, each input unit is connected to each neuron on the map. These connections are not weighted. Instead there is an n-dimensional weight vector $\boldsymbol{w}_i \in V$ attached to each unit on the map layer, with V being a metric space. These vectors are combined to a weight matrix $\boldsymbol{W}$, where the ith row represents the weight vector $\boldsymbol{w}_i$.

4.2 SOM Training

Training of the SOM is performed by iteratively adapting the weights regarding some input data set $\boldsymbol{X} \subseteq V$. The $\boldsymbol{x}_i \in \boldsymbol{X}$ are successively presented to the SOM through the input layer. Given a suitable metric d in V the *best matching unit* (bmu) is determined by

$$c = \arg\min_j \; d(x_i, w_j) \tag{11}$$

for each x_i. The bmu as well as the units in its proximity are updated regarding a time dependent neighborhood function

$$h_{c,i}(t) = \alpha(t) \exp\left(\frac{\|r_c - r_i\|^2}{2\sigma^2(t)}\right), \tag{12}$$

where r_c and r_i are the *coordinates on the map layer* of the bmu and the ith unit. The function smoothes the neighborhood of the bmu. It assures that units are adapted

inversely proportional to their distance to the bmu. α is a time dependent learning rate parameter, which controls how much the map units are affected by the input.

4.3 Clustering by Visualization

There are various ways of visualizing a self-organizing map. In this chapter, we focus on the u-matrix [48], which is a useful tool for clustering via visualization, since it emphasizes cluster borders on the grid. The u-matrix is a two-dimensional array of the same size as the SOM. For each map unit $m_{i,j}$ the distance to the units in its von Neumann neighborhood is calculated. The mean of the results is stored in entry (i, j) of the u-matrix. The u-matrix entries thus represent the average distance of the corresponding map units to their adjacent units. Clusters can be characterized as accumulations of objects with similar features. Given that the features constitute vectors of a metric space, we can interpret the distance between the objects as a measure of their similarity. The SOM models the probability density within the data set. That is, the more densely an area of the feature space is occupied, the more SOM units represent this area. The same holds likewise for sparse areas. SOM weight vectors tend to have small distances if they lie in a cluster and big distances otherwise. Consider, for example, Fig. 4. It displays the u-matrix of a 20 SOM trained on an artificial data set containing three separate Gaussian distributions. The left plot shows the color coded data set with a fitted SOM. Small black dots represent the weight vectors and the connecting lines illustrate the neighborhood relations on the map. The plot on the right hand side is a pseudo-color plot of the u-matrix. Dark blue tones represent small distances between the neighboring units, whereas yellow tones represent big distances and thus cluster borders.

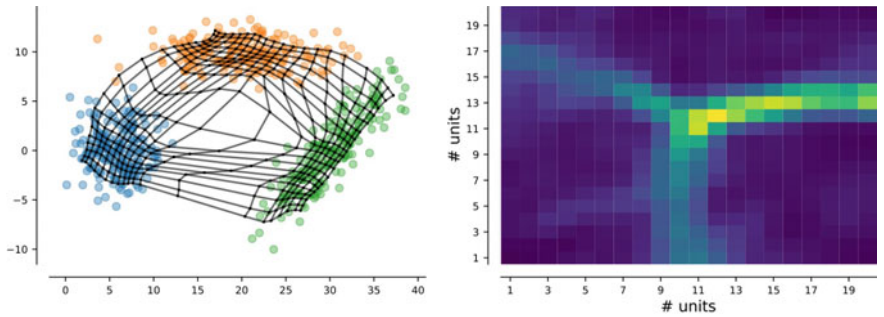


Fig. 4 Demonstration of the u-matrix. Left: Data set sampled from three Gaussian distributions (colored dots) and fitted 20 × 20 SOM (black gird) Right: u-matrix of SOM. Light colors represent big mean distance in the neighborhood, that is, cluster borders

5 Advantages of the Self-organizing Maps

The Self-organizing map has several advantages over other visualization techniques regarding the scope of this work. In comparison to principal component analysis (PCA), SOM performs non-linear dimensionality reduction, while preserving the topology of the original space. PCA reduces the dimensionality linearly by definition, since it is basically a linear transformation.

Manifold methods, like for example, Sammon's mapping, multidimensional scaling (MDS), Isomap, and the well known t-distributed stochastic neighborhood embedding (t-SNE) (to name just a few) have several drawbacks that do not allow their usage in our system. To support ethnomusicologists in formulating new hypotheses, the system has to allow to efficiently compare new material to the existing archive. For example, during a field trip, a musicologist may encounter a rhythm that she needs to check against the archive. Such a workflow can principally not be realized with the mentioned methods, because they compute a low-dimensional embedding of the input data. Hence, comparing new data items to the existing archive would require to retrain the whole system. With the SOM, however, it is possible to predict coordinates in the low-dimensional space for items not involved in the training process.

t-SNE does additionally not preserve the distances in the original. Consider, for example, Fig. 5, which displays t-SNE embeddings of the artificial data set of Fig. 4 for different perplexities. In each embedding one misclassified sample appears in the blue cluster. The embeddings suggest that this sample is far away from the orange cluster and hence that it is very dissimilar to the other members of the orange cluster. Therefore, t-SNE seems not to preserve distances of the original data [49].

Furthermore, techniques like Sammon's mapping and t-SNE are computationally expensive compared to the SOM [49, 50]. The complexity of t-SNE scales quadratically with the number of input samples. Until now, this does not allow to apply the method to large data sets efficiently.

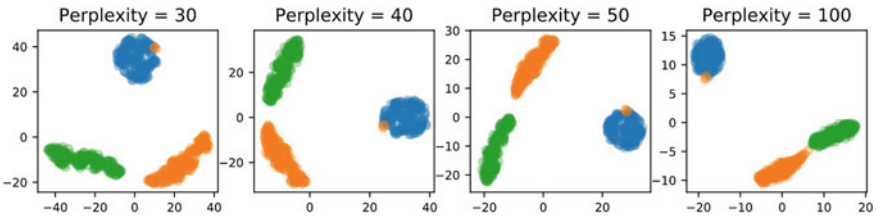


Fig. 5 t-SNE embedding of data set of Fig. 4 for different perplexities. t-SNE does not preserve distance: misclassified sample appears to be far away from its cluster

Table 1 Number of missing values per meta data field in the Wilhelm Heinitz Collection of African Music

Field	Missing values (%)
Country	1
Ethnic group	11
Instrumentation	5
Publisher	2
Recording title	1
Type/Form	9
Year of recording	43

6 The ESRA System

6.1 System Overview

We proposed a system that visualizes the contents of music archives by rhythm similarity as illustrated in Sect. 3. Each recording is marked on a two-dimensional map. Marks that are closer on the map are more similar in terms of the above rhythm definition. There are many similar approaches to automatic organization of music archives. However, they all have certain drawbacks, which do not allow an application on ethnomusicological archives. Our systems tries to overcome this problem. It extracts spectral centroid at each onset from each recording in the archive. In the second, step an HMM is trained on these time series. The HMMs, which represent a high-level rhythm feature in terms of timbre, are automatically organized by a self-organizing map. The system consists of the following processing units.

6.2 Data Set

The Ethnographic Sound Recordings Archive of the University of Hamburg (ESRA) consists of different record collections. In this paper we consider a particular subset of the ESRA, the Wilhelm Heinitz Collection of African Music. This collection is named after musicologist Wilhelm Heinitz (1883–1963), who recorded the contained pieces during filed studies between 1916 and 1948. The pieces are physically available as gramophone records. Each one was converted to 24 bit PCM mono .wav file at a sampling rate of 96 kHz and stored in the ESRA. The pieces are publicly available on the ESRA homepage as mp3 audio streams. The maximum spectral centroid over all files is 3196 Hz. To decrease the computational load, all files were downsampled to 9600 Hz. The collection features 392 recordings of diverse musical content (Fig. 6).

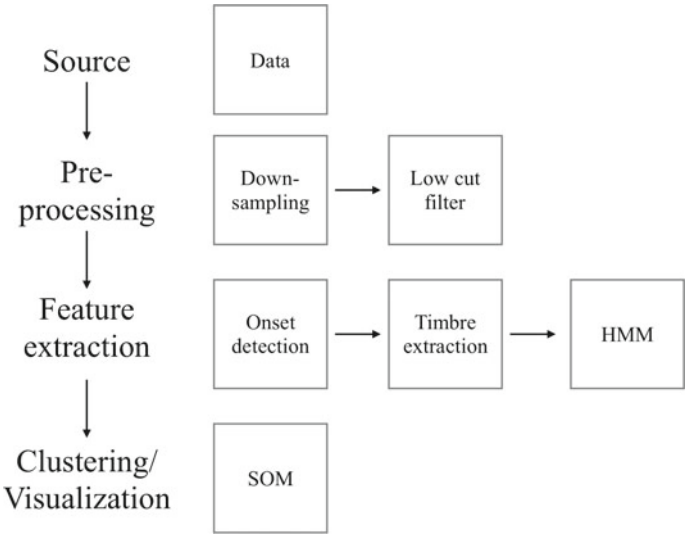


Fig. 6 Overview of the ESRA system

6.3 Preprocessing

The source defines the pipe line which delivers the audio files. In this case it refers to the Wilhelm Heinitz Collection of African Music included in the ESRA. The database engine and the analysis framework are deeply integrated, such that source switching may be accomplished easily. The system is, therefore, portable to any audio collection.

In the preprocessing stage each audio file is resampled in order to reduce computational load. Gramophone recordings have limited frequency range. Analysis of the given collection revealed that the highest spectral centroid is at $f_c^{max} \approx 3196$ Hz. An eighth order Chebyshev type I filter was applied to the records in order to prevent aliasing. Then the pieces were downsampled to $f_s = 9600$ Hz.

6.4 Onset Detection

Onset detection is generally performed by reduction and peak picking. In the first step the input signal is reduced to an onset detection function (ODF), which emphasizes the temporal locations of note onsets. Local maxima of the ODF refer to possible note onsets. Depending on the reduction process and the quality of the audio signal, the maxima of the odf vary considerably in shape and may be masked by noise. Hence, a robust peak picking algorithm has to be applied.

The choice of the reduction algorithm depends on the content of the audio signal. Bello et al. [51] reported that time domain methods are especially suitable for percussive music, whereas spectral methods perform better on music with pitched instruments. Dixon [52] reviewed different approaches to onset detection. Evaluation revealed that onset detection functions based on spectral flux (SF) and complex domain (CD) functions perform better with phase based methods. Moreover, there is not a significant difference in their results. He argues that slight differences originate from implementation details and parameter settings. He hence concludes that the choice of algorithm may be based on factors unrelated to the detection performance.

From this background we choose to use the spectral flux algorithm:

$$SF(n) = \|\mathrm{H}\left(\boldsymbol{X}(n) - \boldsymbol{X}(n-1)\right)\| \quad . \tag{13}$$

Spectral flux refers to the bandwise difference in energy between the spectra of two consecutive time windows $\boldsymbol{X(n)}$ and $\boldsymbol{X(n-1)}$, where n is an integer index. Since note onsets mostly coincide with an increase in energy, the half-rectifier window

$$\mathrm{H}(x) = \frac{x + |x|}{2} \tag{14}$$

is applied in order to only capture positive shifts in spectral energy. Several spectral flux definitions have been proposed, which differ regarding the choice of norm and the application of the half-rectifier. In this paper we follow the approach of [26], who introduced a per window scaling of the SF by the magnitude of the current window:

$$SF'(t) = \frac{\sum_n \mathrm{H}\left(|\boldsymbol{X}(t,n)| - |\boldsymbol{X}(t,n-1)|\right)}{\|\boldsymbol{X}(t,n)\|} \quad . \tag{15}$$

This has the effect of eliminating performance dynamics such that the algorithm emphasize changes in timbre instead of loudness. We additionally apply a low pass filter with a Hamming window. Window length of 5 points performed best given the data sets. Three mentioned detection functions are displayed in Fig. 7. The first plot displays the time domain signal under consideration, female solo vocals with accompanying claves [53].

The proposed algorithm has been evaluated on the dataset provided by [54, 55]. Evaluation results are reported in Table 2.

In the case of ESRA the Spectral Flux detection function seems to be an adequate choice since the algorithm is easy to implement and has low computational effort.

Table 2 Evaluation of the onsets detection algorithm proposed in this paper

Data set	Precision	Recall	F-measure
Glover et al.	0.54	0.78	0.59 ± 0.326
Leveau et al.	0.73	0.86	0.77 ± 0.187

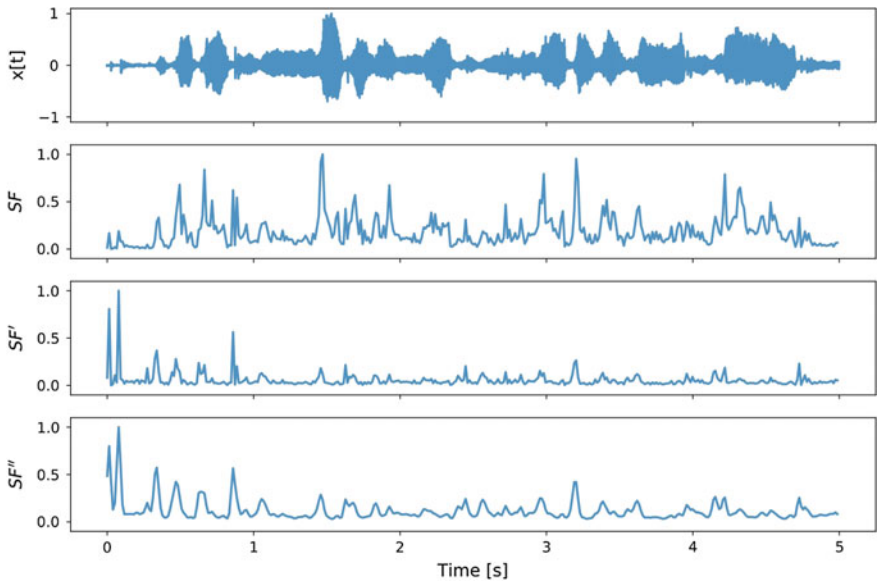


Fig. 7 Comparison of three different definition of Spectral Flux. (1) Time domain signal. (2) Spectral Flux (3) Spectral Flux with magnitude based scaling (4) Scaled Spectral Flux and additional smoothing with 5 point Hamming kernel

6.5 Feature Extraction

Spectral centroid is extracted as a scalar model of timbre. It is calculated at each onset from a 512 point (≈50 ms) discrete Fourier Transform using the Hamming window function. A HMM is trained on each record's SC time series with $m = 4$. The vector of state-dependent means $\boldsymbol{\lambda}$ was initialized as the mth percentile of the respective training data to assure an equidistant distribution of the initial guesses. The transition probability matrix $\boldsymbol{\Gamma}$ is initialized with the most probability mass on the main diagonal and linearly decreasing mass on the other transitions. This alleviates the training process to converge to a pseudo forward model, in which most probable hidden state sequence loops from 1 to m. We favor this type of model, since it would represent a recurring rhythm pattern.

6.6 Feature Selection

Examination of the state dependent distributions revealed high correlations between the means (see Table 3). For that reason, we decided to exclude them from the SOM training and to only use the transition probability matrices.

Table 3 Correlation of the state dependent means of each HMM in the training set

λ_1	λ_2	λ_3	λ_4
1	0.873	0.755	0.657
	1	0.909	0.818
		1	0.928
			1

6.7 Self-organizing Map

The COMSAR project is currently in the prototype phase. At this stage, all SOMs have rectangular grid topology. The map unit's weight vectors are initialized by random variates of m Dirichlet distributions, one for each state. The α parameter is set such that the mth distribution puts 50% of the probability mass to the mth vector component. This procedure samples probability matrices with high values on the main diagonal. We utilize a normal distributed neighborhood kernel. Neighborhood radius decreases linearly to 1. We assume the transition probability matrices to be a rhythm finger print and use only this part of the HMMs as input to the SOM. Distances in the feature space are measured using squared euclidean distance. We further use the Batch-training algorithm, which updates the weight vectors after a presentation of the full data set. This kind of algorithm is especially successful given roughly ordered weight vectors.

7 Experiment

7.1 Setup

For this example, we trained a self-organizing map of 1600 units, regularly arranged on 40×40 grid. The neighborhood size was initialized to 30 and allowed to decrease linearly to 1. The Batch-Map was applied in order to avoid specifying a learning rate factor. The number of iterations was set to 1000. Commonly, SOMs have to be trained with much more iterations to reach an equilibrium state. However, the global ordering of the SOM is usually established after a few hundred iterations. We used 389 records of the Wilhelm Heinitz Collection included in ESRA as training set. 4-state HMMs were fitted to each one. Four models failed due to too short training sequences. These were excluded from the further procedure.

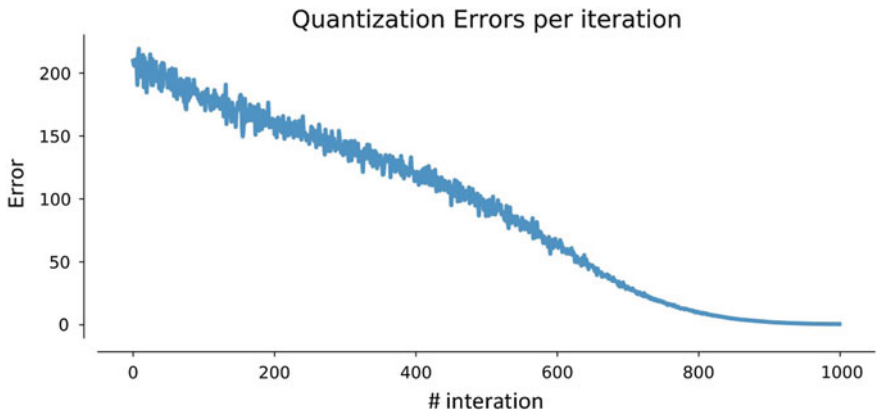


Fig. 8 The quantization per iteration

7.2 Result

The goal of exploratory data analysis is to develop hypotheses about the data from visualizations. Such tasks typically employ unsupervised methods, which usually do not rely on target values. Since there is no ground truth for the data set, it is not easy to quantify the performance of the system. There are several measures, which all shed some light on different aspects of the SOM training as well as phenomenons of similarity computations in high-dimensional spaces in general [56, 57]. An exhaustive presentation of such measure goes beyond the scope of this chapter. We hence provide the most typical measure, the quantization error (QE). The QE illustrates how well the SOM adapts to the feature space by summing the differences between the inputs and their bmus. In Fig. 8 the QE is depicted for each iteration. The initial drop represents a sudden ordering on global scale. The smooth development of the error is typical for the Batch-Map. Since the error tends to zero we assume that the SOM models the feature space well.

To asses the applicability of the system for ethnomusicological research, we have to await feedback form expert users. Appropriate surveys will be carried out during the evaluation phase of the project and are hence not available at this time. We, therefore, showcase a qualitative examination of the u-matrix by means of single examples. The reader is encouraged to reenact these examples by listening to the audio recordings linked in the following.

Figure 9 displays the u-matrix of the trained SOM. Dark blue regions refer to coarse clusters, whereas lighter colors represent cluster borders. Red crosses mark the position of the best matching units for each example of the training data. A clustering structure is clearly visible. The corners of the SOM seem to represent highly diverse features. Listening to the corresponding records revealed that these are mostly test tones. Test tones are 8–12 s long records of a steady reed instrument tone. This suggests that a toroid or even spherical shaped map layer is likely to be

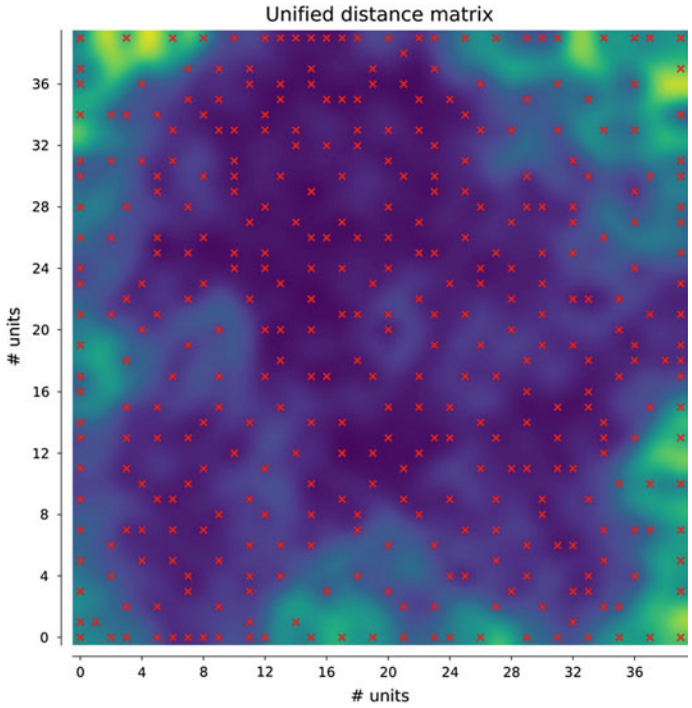


Fig. 9 U-matrix: dark tones represent cluster, lighter tones cluster borders. Red crosses mark the positions of the best matching units of the input data

preferred over the rectangular grid. These shapes connect opposing borders of the map and thus help to avoid border effects.

How features of records in the corners of the SOM differ from others can be seen in Fig. 10. It displays the weight vectors of the upper left 12 by 12 area as time series. Note how the SOM achieves an ordering of the features that is immediately plausible to the human viewer. Models in the corners tend to have their probability mass centered on the transitions from state 1 to state 4 and state 2 to state 4. The more one moves to the center of the SOM, the more the probability is distributed evenly over all transitions. Further analysis showed that even though the probabilities are distributed evenly, they still favor a forward-model style, where the major transition probability mass is on the main diagonal.

In the neighborhood of neuron $(20, 12)$, records are gathered that feature a regularly played percussion instrument (EAn84-B1,[6] EAn86-A2,[7] EAn35-A1,[8]

[6] http://esra.fbkultur.uni-hamburg.de/explore/view?entity_id=216.

[7] http://esra.fbkultur.uni-hamburg.de/explore/view?entity_id=219.

[8] http://esra.fbkultur.uni-hamburg.de/explore/view?entity_id=75.

Fig. 10 Lines plots of the weight vectors of the 12 × 12 units from the upper left area

EAn101-A1[9]). Especially EAn101-A1 is an interesting example, since it also features a strong beat.

The influence of the features on the units can be depicted more generally by component planes. Figure 11 displays how the feature contributes to each position. Again, it can be seen that most activation is on the main diagonal, which indicates forward-models, which means that the most probable hidden sequence of the HMM tend to be loops.

A further approach to SOM evaluation is through map calibration. To each unit a label is assigned that represents the class membership of its best matching training example according to a given set of meta data. If the SOM ordered the features according to the given meta data, cluster structures emerge in the visualization. Figure 12 illustrates calibrations to *ethnic group*, *instrumentation*, *type/form*, *country*, *year of recording* and *publisher*. Each color represents a class of the meta data

[9] http://esra.fbkultur.uni-hamburg.de/explore/view?entity_id=261.

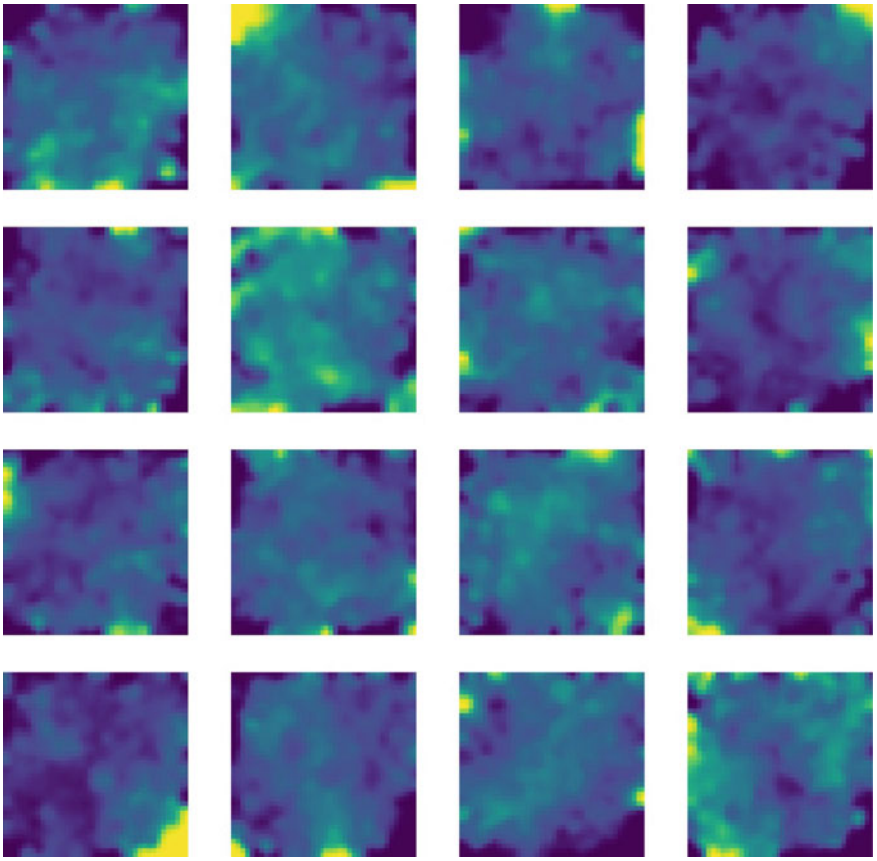


Fig. 11 Component planes illustrate the contribution of each feature per map unit. Lighter tones represent high contribution. The plot represents the features as they are ordered in the HMM. The plane in the upper left corresponds to the transition probability from state 1 to state 1. The next plot to the right corresponds to the transition from state 1 to state 2, etc.

set. Some nuances are hardly visible because of the number of classes. Country, for example, features 40 classes. Feasible clusterings are only archived by *year of recording* and *publisher*. However, this needs further investigation. It is possible that the clustering only appears to be better on the first glance, because the ration of the number of classes to the number of map units is much better. *year of recording* has 9 classes, *publisher* 10. Obviously the SOM does not cluster according to any available meta data.

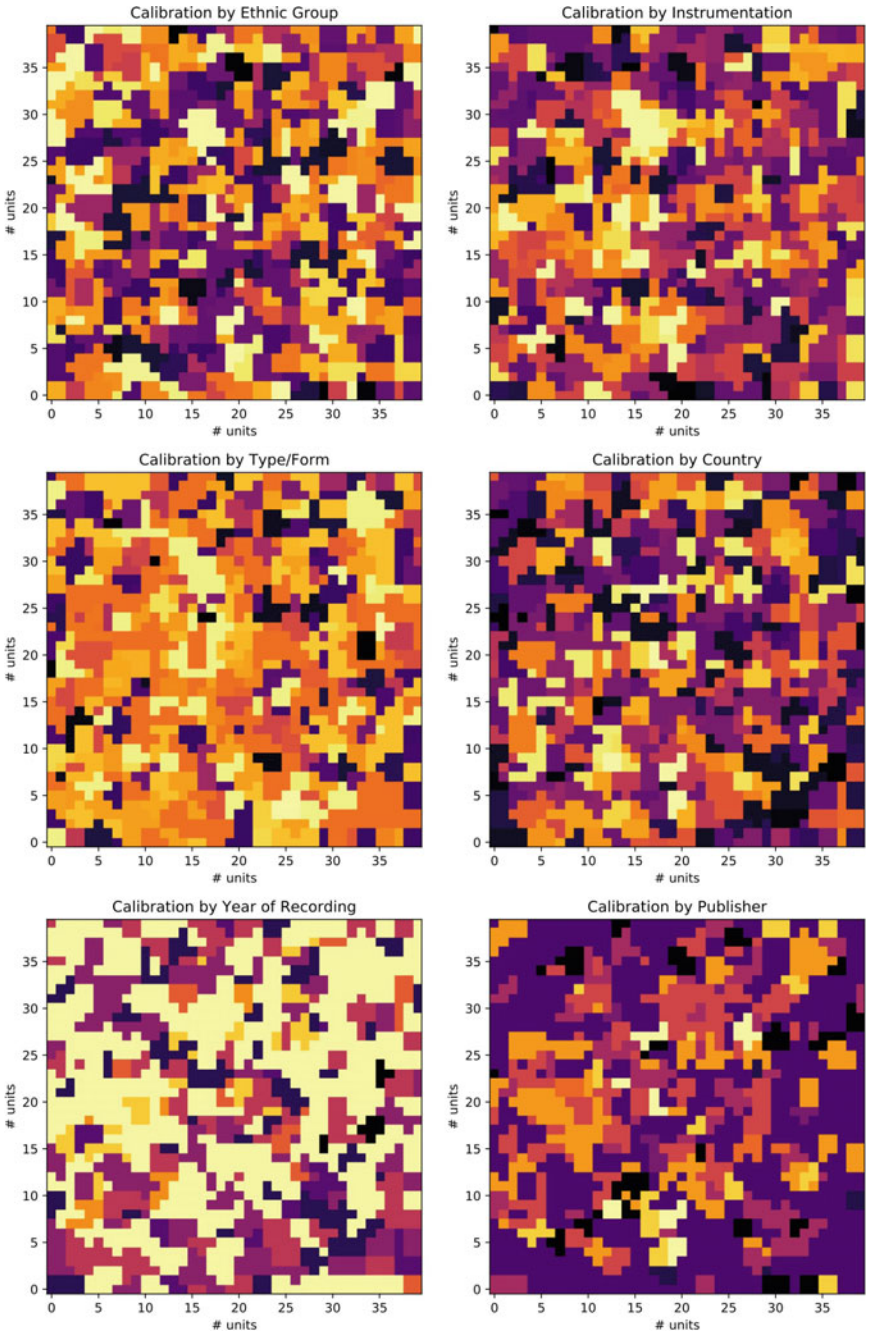


Fig. 12 SOM calibration given diverse sets of meta data

8 Conclusion

In this paper we proposed a system for the exploration of ethnographic music archives. We reviewed the literature on automatic organization systems for music collections and discussed how they could be applied to ethnographic music archives, which exhibit unique characteristics. We gave an introduction to the core algorithms of the COMSAR project as well as a complete description of the modeling process. An example application of the system demonstrated possible evaluation strategies. It was shown that the system does not cluster according to the meta data fields available in our data set (*ethnic group*, *instrumentation*, *type/form*, *country*, *year of recording* and *publisher*). This was expected, since particular timbre sequences are not restricted to a country or a musical form. Furthermore, an example was provided that illustrated how musical patters may be clustered. Further evaluation is only possible with skilled expert users. Adequate studies will be carried out in the evaluation phase of the COMSAR project.

However, empirical results show that alternative SOM topologies may help to improve the clustering ability. Furthermore, more flexible algorithms like the growing self-organizing map (GSOM) and Neural Gas need to be tested. Additionally, the squared euclidean distance may not be an optimal metric for the given data, since the subspace in which m-dimensional stochastic matrices reside, differs from $\mathbb{R}^m$. Hence, this system is an example of dimensionality reduction on the simplex. Alternative methods have to be tested that measure distances between probability distributions, for example the Hellinger and the Aitchison distances [58].

References

1. van Kranenburg P, de Bruin M, Volk A (2017) Documenting a song culture: the Dutch song database as a resource for musicological research. Int J Digit Libr 1–11
2. Fillon T, Simonnot J, Mifune M-F, Khoury S, Pellerin G, Coz ML, de la Bretèque EA, Doukhan D, Fourer D (2014) Telemeta: an open-source web framework for ethnomusicological audio archives management and automatic analysis. In: Proceedings of the 1st international workshop on digital libraries for musicology, New York, pp 1–8
3. Abdallah S, Benetos E, Gold N, Hargreaves S, Weyde T, Wolff D (2017) The digital music lab: a big data infrastructure for digital musicology. ACM J Comput Cult Herit 10(1)
4. Pfeiffer S, Fischer S, Effelsberg W (1996) Automatic audio content analysis. In: Proceedings of the forth ACM international conference on multimedia, Boston, MA, USA, November 1996
5. Melucci M, Orio N (1999) Music information retrieval using melodic surface. In: Proceedings of the fourth ACM conference on digital libraries, Berkley, CA, USA, August 1999, pp 152–160
6. Tseng Y-H (1999) Content-based retrieval for music collections. In: Proceedings of the 22nd annual international ACM SIGIR, Berkeley, CA, USA, August 1999, pp 176–182
7. Melucci M, Orio N (2000) Smile: a system for content-based music information retrieval environments. In: RIAO' 2000 conference proceedings, vol 2, pp 1261–1275
8. Frühwirth M, Rauber A (2001) Self-organizing maps for content-based music clustering. In: Tagliaferri R, Marinaro M (eds) Proceedings of the 12th Italian workshop on neural nets. Perspectives in neural computing, Vietri sil Mare, Salerno, Italy, May 2001

9. Rauber A, Frühwirth M (2001) Automatically analyzing and organizing music archives. In: Constantopoulos P, Sølvberg IT (eds) Research and advanced technology for digital libraries. Lecture notes in computer science, Darmstadt, September 2001, pp 402–414
10. Pamplak E (2001) Islands of music. PhD dissertation, Institut für Softwaretechnik und Interaktive Systeme der Technischen Universit at Wien, Dezember 2001
11. Juhász Z (2009) Automatic segmentation and comparative study of motives in eleven folk song collections using self-organizing maps and multidimensional mapping. J New Music Res 38(1):77–85
12. Juhász Z (2011) Low dimensional visualization of folk music systems using the self organizing cloud. In: Klapuri A, Leider C (eds) Proceedings of the 12th international society for music information retrieval conference, ISMIR 2011, Miami, Florida, USA, 24–28 October 2011. University of Miami, pp 299–304 [Online]. http://ismir2011.ismir.net/papers/OS3-2.pdf
13. Panteli M, Benetos E, Dixon S (2016) Learning a features space for similarity in world music. In: Proceedings of the 17th international society for music information retrieval conference
14. Al Mansouria HZ. Helv el mabassem. http://esra.fbkultur.uni-hamburg.de/explore/view?entity_id=514
15. Mohamed Eff. el Akkad C. Taxim rast (ala alwahda). http://esra.fbkultur.uni-hamburg.de/explore/view?entity_id=514
16. Blaß M (2013) Timbre-based rhythm theory using Hidden Markov models. Master's thesis, University of Hamburg
17. Blaß M (2013) Timbre-based drum pattern classification using Hidden Markov models. In: Proceedings of the 6th international workshop on machine learning and music, ECML/PKDD
18. Mauch M, Dixon S (2012) A corpus-based study of rhythm patterns. In: Proceedings of the 13th international society for music information retrieval conference (ISMIR)
19. Desain P (1992) A (de)composable theory of rhythm perception. Music Percept 9(4):439–454
20. Alluri V, Toiviainen P (2009) Exploring perceptual and acoustical correlates of polyphonic timbre. Music Percept Interdiscip J 27(3):223–242
21. Zucchini W, MacDonald IL (2009) Hidden Markov models for time series. Monographs on statistics and applied probability, vol 110. Chapman & Hall, Boca Raton
22. Aucouturier J-J, Sandler M (2001) Segmentation of musical signals using Hidden Markov models. In: Proceedings of the 110th audio engineering society, Amsterdam, The Netherlands, May 2001
23. Mavromatis P (2012) Exploring the rhythm of the palestrine style: a case study in probabilistic grammar induction. J Music Theory 56(2):169–223
24. Shao X, Xu C, Kankanhalli M (2004) Unsupervised classification of music genre using hidden Markov model. In: IEEE international conference on multimedia and expo (ICME), vol 3, pp 2023–2026
25. Braasch J (2013) The $\mu\cdot$ cos*m* project: an introspective platform to study intelligent agents in the context of music ensemble improvisation. In: Bader R (ed) Sound–perception–performance. Current research in systematic musicology, vol 1. Springer, Heidelberg
26. Alexandraki C (2014) Real-time machine listening and segmental re-synthesis for networked music performance. PhD dissertation, University of Hamburg
27. Rabiner LR, Juang BH (1986) An introduction to Hidden Markov models. IEEE ASSP Mag
28. Rabiner LR (1989) A tutorial on Hidden Markov models and selected applications in speech recognition. In: Proceedings of the IEEE, vol 77, no 2. IEEE, pp 257–286
29. Aucouturier J-J, Pachet F (2002) Music similarity measures: what's the use? In: Proceedings of the 3rd international society for music information retrieval conference, ISMIR
30. Aucouturier J-J, Pachet F, Sandler M (2005) The way it sounds: timbre models for analysis and retrieval of music signals. IEEE Trans Multimed 7(6):1028–1035
31. Davis S, Mermelstein P (1980) Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. IEEE Trans Acoust Speech Signal Process 28(4):357–366
32. Iverson P, Krumhansl CL (1993) Isolating the dynamic attributes of musical timbre. J Acoust Soc Am 94(5):2595–2603

33. McAdams S, Winsberg S, Donnadieu S, Soete GD, Krimphoff J (1995) Perceptual scaling of sythesized musical timbres: common dimensions, specificities, and latent subject classes. Psychol Rev 58:177–192
34. Hourdin C, Charbonneau G, Moussa T (1997) A multidimensional scaling analysis of musical instruments' time-varying spectra. Comput Music J 21(2):44–55
35. von Bismarck G (1974) Timbre of steady sounds: a factorial investigation of its verbal attributes. Acoustica 3(3):146–159
36. Zacharakis AI, Pastiadis K, Papadelis G, Reiss JD (2011) An investigation of musical timbre: uncovering salient semantic descriptors and perceptual dimensions. In: Klapuri A, Leider C (eds) Proceedings of the 12th international society for music information retrieval conference, ISMIR 2011, Miami, Florida, USA, 24–28 October 2011. University of Miami, pp 807–812 [Online]. http://ismir2011.ismir.net/papers/OS10-3.pdf
37. Grey JM (1977) Multidimensional perceptual scalings of musical timbres. J Acoust Soc Am 61(5):1270–1277
38. Grey JM, Gordon JW (1978) Perceptual effects of spectral modifications on musical timbres. J Acoust Soc Am 63(5)
39. Schubert E, Wolfe J, Tarnopolsky A (2004) Spectral centroid and timbre in complex, multiple instrumental textures. In: Proceedings of the 8th international conference on music perception and cognition, pp 654–657
40. Schubert E, Wolfe J (2006) Does timbral brightness scale with frequency and spectral centroid. Acta Acoust 92(2):820–825
41. Siedenburg K, Fujinaga I, McAdams S (2016) A comparison of approaches to timbre descriptors in music information retrieval and music psychology. J New Music Res 45(1):27–41
42. Park Y-S, Chon T-S, Bae M-J, Kim D-H, Lek S (2017) Ecological informatics. In: Multivariate data analysis by means of self-organizing maps. Springer, pp 251–272
43. Resta M (2014) Financial self-organizing maps. In: Proceedings of the 24th international conference on artificial neural networks, Hamburg, pp 781–788
44. Toiviainen P (2005) Visualization of tonal content with self-organizing maps and self-similarity matrices. ACM Comput Entertain 3(4):1–10
45. Vembu S, Baumann S (2004) A self-organizing map based knowledge discovery for music recommendation systems. In: Computer music modeling and retrieval: second international symposium (CMMR), vol 3310. Lecture notes in computer science, Esbjerg, Denmark, May 2004
46. Ness SR, Tzanetakis G (2009) Somba: multiuser music creation using self-organizing maps and motion tracking. In: Proceedings of the international computer music conference (ICMC)
47. Odowichuk G, Tzanetakis G (2012) Browsing music in and sound using gestures in a self-organized 3d space. In: Proceedings of the international computer music conference (ICMC)
48. Lötsch J, Ultsch A (2014) Exploiting the structures of the u-matrix. In: Proceedings of the 10th international workshop on self-organizing maps, pp 249–257
49. van der Maaten L, Hinton G (2008) Visualizing data using t-SNE. J Mach Learn Res 9:2579–2605
50. Flexer A (2001) On the use of self-organizing maps for clustering and visualization. Intell Data Anal 1:373–384
51. Bello JP, Daudet L, Abdallah S, Duxbury C, Davis M, Sandler MB (2005) A tutorial on onset detection in music signals. IEEE Trans Speech Audio Process 13(5):1035–1047
52. Dixon S (2006) Onset detection revisited. In: Proceedings of the 9th international conference on digital audio effects (DAFx-06), pp 18–20
53. n'Dri L, Aya T, n'Dri Akissi K. Aoussi. http://esra.fbkultur.uni-hamburg.de/explore/view?entity_id=514
54. Glover J, Lazzarini V, Timoney J (2011) Real-time detection of musical onsets with linear prediction and sinusoidal modelling. J Adv Signal Process 68:297–316
55. Leveau P, Daudet L, Richard G (2004) Methodology and tools for the evaluation of automatic onset detection algorithms in music. In: Proceedings of the 5th international conference on music information retrieval

56. Flexer A, Schnitzer D, Schlüter J (2012) A MIREX meta-analysis of hubness in audio music similarity. In: Proceedings of the international conference on music information retrieval
57. Flexer A (2015) Improving visualization for high-dimensional music similarity spaces. In: Proceedings of the 16th international conference for music information retrieval
58. Le T, Cuturi M (2015) Unsupervised Riemannian metric learning for histograms using Aitchison transformations. In: Proceedings of the 32nd international conference on machine learning, vol 37
59. Klapuri A, Leider C (eds) (2011) Proceedings of the 12th international society for music information retrieval conference, ISMIR 2011, Miami, Florida, USA, 24–28 October 2011. University of Miami [Online]. http://ismir2011.ismir.net/

Spatial Manipulation of Musical Sound: Informed Source Separation and Respatialization

Sylvain Marchand

Abstract "Active listening" enables the listener to interact with the sound while it is played, like composers of electroacoustic music. The main manipulation of the musical scene is (re)spatialization: moving sound sources in space. This is equivalent to source separation. Indeed, moving all the sources of the scene but one away from the listener separates that source. And moving separate sources then rendering from them the corresponding scene (spatial image) is easy. Allowing this spatial interaction/source separation from fixed musical pieces with a sufficient quality is a (too) challenging task for classic approaches, since it requires an analysis of the scene with inevitable (and often unacceptable) estimation errors. Thus we introduced the informed approach, which consists in inaudibly embedding some additional information. This information, which is coded with a minimal rate, aims at increasing the precision of the analysis/separation. Thus, the informed approach relies on both estimation and information theories. During the DReaM project, several informed source separation (ISS) methods were proposed. Among the best methods is the one based on spatial filtering (beamforming), with the spectral envelopes of the sources (perceptively coded) as additional information. More precisely, the proposed method is realized in an encoder-decoder framework. At the encoder, the spectral envelopes of the (known) original sources are extracted, their frequency resolution is adapted to the critical bands, and their magnitude is logarithmically quantized. These envelopes are then passed on to the decoder with the stereo mixture. At the decoder, the mixture signal is decomposed by time-frequency selective spatial filtering guided by a source activity index, derived from the spectral envelope values. The real-time manipulation of the sound sources is then possible, from musical pieces initially fixed (possibly on some media like CDs), and with an unpreceded (controllable) quality.

S. Marchand (✉)
University of La Rochelle, La Rochelle, France
e-mail: sylvain.marchand@univ-lr.fr

R. Bader (ed.), *Computational Phonogram Archiving*, Current Research in Systematic Musicology 5, https://doi.org/10.1007/978-3-030-02695-0_8

1 Introduction

Active listening of music is an artistic as well as a technological topic of growing interest, that concerns offering listeners the possibility to interact in real time with the music, e.g. to modify the elements, the sound characteristics, and the structure of the music while it is played. This involves, among other examples, advanced remixing processes such as generalized karaoke (muting any musical element, not only the lead vocal track), respatialization, or upmixing. The applications are numerous, from learning/teaching of music to gaming, through new creative processes (disc jockeys, live performers, etc.). In the context of ethnomusicological archiving, the recordings can consist of several tracks, but for the purpose of compatibility, only the mix can often be distributed in the archive. Thus, a technique allowing the user to get access back to the separate tracks from the stereo mix can be very useful.

To get this new freedom, a simple solution would be to give the user access to the individual tracks that compose the mix [24], by storing them into some multitrack format. This approach has two main drawbacks: First, it leads to larger multitrack files. Second, it yields files that are not compatible with the prevailing stereo standards.

Another solution is to perform some blind separation of the sources from the stereo mix. The problem is that even with state-of-the-art blind source separation techniques the quality is usually insufficient (estimation errors are unavoidable, e.g. see [2]) and the computation is heavy [1, 18].

In the DReaM project [16], we proposed a system designed to perform source separation and accurately recover the separate tracks from the stereo mix. The system consists of a coder and a decoder.

The coder is used at the mixing stage, where the separate tracks are known. It determines the information necessary to recover the tracks from the mix and embeds it in the mix. In the classic case of Pulse-Code Modulation (PCM), this information is inaudibly hidden in the mix by a watermarking technique [22]. In the case of compressed audio formats, it can be embedded in a dedicated data channel or directly in the audio bitstream. With a legacy system, the coded stereo mix can be played and sounds just like the original, although it includes some additional information. Apart from backward compatibility with legacy systems, a further advantage concerns the fact that the file size stays comparable to the one of the original mix, since the additional information sent to the decoder is rather negligible.

This decoder performs source separation of the mix with parameters given by the additional information. This Informed Source Separation (ISS) approach [10] permits producing good separate tracks, thus enabling active listening applications.

The original target of the DReaM project was the music industry, which turned out to be quite conservative. For instance, they appeared to be reserved with the use of audio formats that are alternative to conventional stereo encoding, hence hindering the development of object-based formats or advanced spatial audio formats such as Ambisonics. Another example is the fact that listeners are considered as

(passive) consumers, even if some want to behave as musicians (active listeners, content producers, etc.).

Fortunately, there is some opportunity for the system developed in the project for an application for musical archives. Indeed, some recordings contain several tracks, but the diffusion format is still legacy stereo. Thus, having a format backward compatible with standard stereo but allowing to recover the individual tracks present in the mix can be of interest. The DReaM project showed that it is possible.

The chapter is organized as follows. Section 2 presents the DReaM project: its fundamentals and target applications. Section 3 introduces the mixing models we are considering, Sect. 4 describes the separation/unmixing methods developed in the project, and Sect. 5 illustrates the working prototypes available for demonstration purposes. Finally, Sect. 6 draws some conclusions.

2 The DReaM Project

DReaM[1] is a French acronym for "*le Disque Repensé pour l'écoute active de la Musique*", which means "the disc thought over for active listening of music". This is the name of an academic project with industrial finality, coordinated by the author, and funded by the French National Research Agency (ANR). The project involved academic partners (LaBRI—University of Bordeaux, Lab-STICC—University of Brest, GIPSA-Lab—Grenoble INP, LTCI—Telecom ParisTech, ESPCI—Institute Langevin) together with iKlax Media, a company for interactive music that contributed to the Interactive Music Application Format (IMAF) standard [9].

The origin of the project comes from the observation of artistic practices. More precisely, composers of acousmatic music conduct different stages through the composition process, from sound recording (usually stereophonic) to diffusion (multiphonic). During live interpretation, they interfere decisively on spatialization and coloration of pre-recorded sonorities. For this purpose, the musicians generally use a mixing console to upmix the musical piece being played from an audio CD. This requires some skills, and imposes musical constraints on the piece. Ideally, the individual tracks should remain separate. However, this multi-track approach is hardly feasible with a typical (stereophonic) audio CD.

Nowadays, the audience is more eager to interact with the musical sound. Indeed, more and more commercial CDs come with several versions of the same musical piece. Some are instrumental versions (e.g. for karaoke), other are remixes. The karaoke phenomenon gets generalized from voice to instruments, in musical video games such as *Rock Band*.[2] But in this case, enabling interaction translates to users having to buy a video game, which includes the multi-track recording.

[1] See URL: http://dream.labri.fr.

[2] See URL: http://www.rockband.com.

Yet, the music industry seems to be reluctant to releasing the multi-track versions of big-selling hits. The only thing the user can get is a standard CD, thus a stereo mix, or its digital version available for download or streaming.

2.1 Project Goals and Objectives

In general, the project aims at solving a so-called inverse problem, to some quality extent, at the expense of additional information. In particular, an example of such an inverse problem can be source separation: recovering the individual source tracks from the given mix.

On the one hand coding the solution (e.g., the individual tracks and the way to combine them) can bring high quality, but with a potentially large file size, and a format not compatible with existing stereo formats.

On the other hand the blind approach (without information) can produce some results, but of insufficient quality for demanding applications (explained below). Indeed, the mixture signals should be realistic music pieces, ideally of professional quality, and the separation should be processed in real-time with reasonable computation costs, so that real-time sound manipulation and remixing can follow. The blind approach can be regarded as an estimation without information, while coding can be regarded as using information (from each source) without any estimation (from the mix).

The informed approach proposed by DReaM is just in between these two extremes: getting musically acceptable results with a reasonable amount of additional information. The problem is now to identify and encode efficiently this additional information [19]. Remarkably, ISS can thus be seen both as a multi-track audio coding scheme using source separation, or as a source separation system helped by audio coding.

This approach addresses the source separation problem in a coder/decoder configuration. At the coder (see Fig. 1), the extra information is estimated from the original source signals before the mixing process and is inaudibly embedded into the final mix. At the decoder (see Fig. 2), this information is extracted from the mix and used to assist the separation process. The residuals can be coded as well, even if joint coding is more efficient (not on the figures for the sake of simplicity, see Sect. 4 instead).

So, a solution can be found to any problem, thanks to the additional information embedded in the mix.

"There's not a problem that I can't fix,
'cause I can do it in the mix!"
(Indeep – Last Night a DJ Saved my Life)

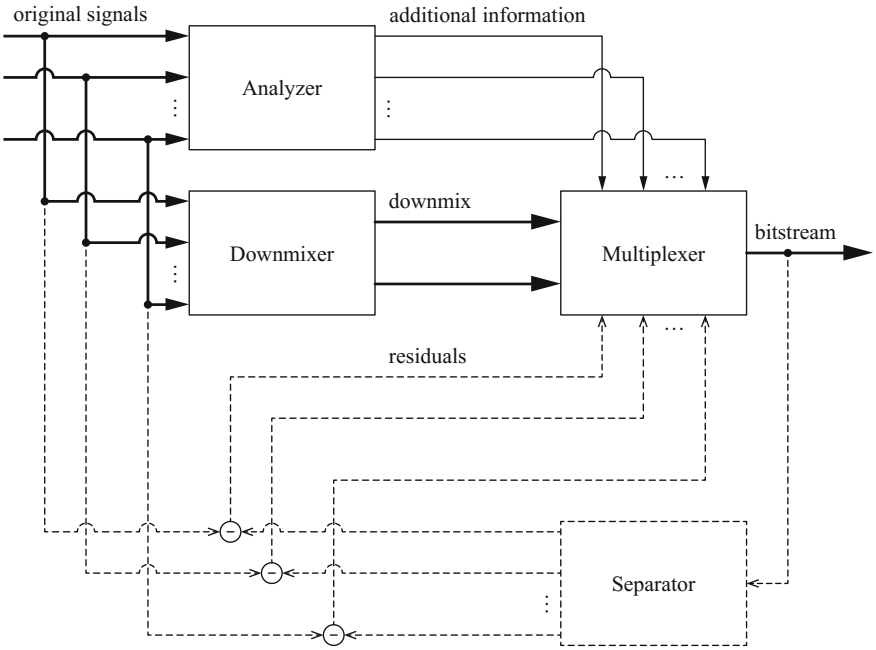


Fig. 1 General architecture of an ISS coder

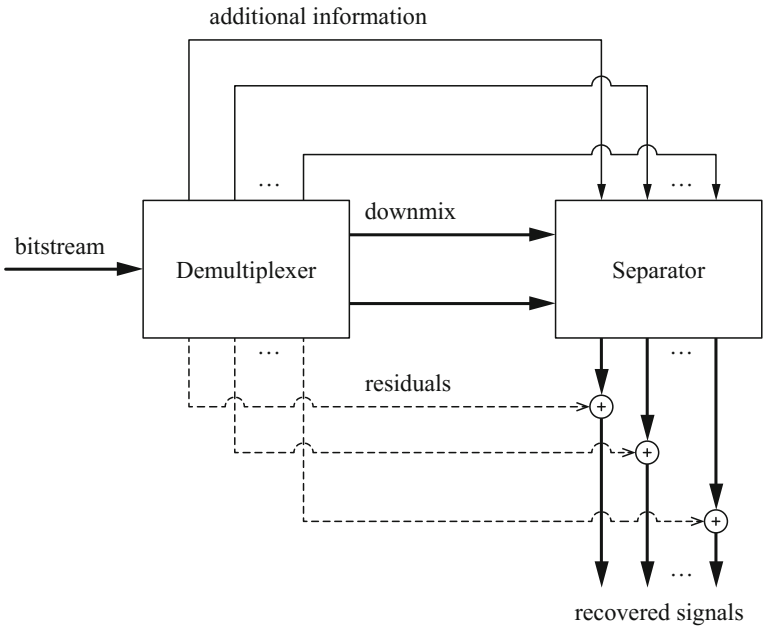


Fig. 2 General architecture of an ISS decoder

2.1.1 From Active Audio CD…

The original goal of the project was to propose a fully backward-compatible audio-CD permitting musical interaction.

The idea was to inaudibly embed (using a high-capacity watermarking technique, see [22]) in the audio track some information enabling to some extent the musical decomposition, that is the inversion of the music production chain: dynamics decompression, source separation (unmixing), deconvolution, etc.

With a standard CD player, one would listen to the fixed mix. With an active player however, one could modify the elements and the structure of the audio signal while listening to the music piece.

2.1.2 …Towards Enhanced Compressed Mix

Now that the music is getting all digital, the consumer gets access to audio files instead of physical media. Although the previous strategy also applies to the (PCM) files extracted from the audio CD, most audio files are distributed in lossy compressed formats (e.g. AAC, MP3, or OGG).

Fortunately, the proposed approach also applies to compressed mixes (see [3, 6]). The extra information can then either be included in some ancillary data, or be embedded (almost) inaudibly in the audio bitstream itself. The latter option is much more complicated, since lossy but perceptually loss-less coding aims at removing inaudible information. Both coders (perceptual and informed) have then to be merged, to maintain a certain information trade-off.

2.2 Applications

Active listening [11] amounts to performing various operations that modify the elements and structure of the music signal during the playback of a piece.

This process, often simplistically called remixing, includes generalized karaoke, respatialization, or applying certain effects to individual audio tracks (e.g., adding some distortion to an acoustic guitar).

The goal is to enable the listener to enjoy freedom and personalizing of the musical piece through various reorchestration techniques.

Alternatively, active listening solutions intrinsically provide simple frameworks to the artists to produce different versions of a given piece of music. Moreover, it is an interesting framework for music learning/teaching applications.

2.2.1 Respatialization

The original application was to let the public experience the freedom of composers of electroacoustic music during their live performances: moving the sound sources in the acoustic space. Although changing the acoustical scene by means of respatialization is a classic feature of contemporary art (electroacoustic music), and efforts have been made in computer music to bring this practice to a broader audience [20], the public seems just unaware of this possibility and rather considered as passive consumers by the music industry. However, during the public demonstrations of the DReaM project, we felt that the public was very reactive to this new way of interacting with music, to personalize it, and was ready to adopt active listening, mostly through musical games.

2.2.2 Generalized Karaoke

Games, or "serious" games, can be very useful for music learning/teaching applications. The generalized karaoke application is the ability to suppress any audio source, either the voice (classic karaoke) or any instrument ("music minus one"). The user can then practice singing or playing an instrument while being integrated in the original mix and not a cover song.

Note that these two applications (respatialization and generalized karaoke) are related, since moving a source far away from the listener will result in its muting, and reciprocally the ability to mute sources can lead to the monophonic case (the spatial image of a single source isolated) where respatialization is much easier (possible to some extent even without recovering the audio object from this spatial image).

2.2.3 Sound Archives

It turns out that the system developed in the project might be very useful for musical archives. Indeed, some recordings contain several tracks, but the diffusion format is still legacy stereo. Thus, having a format backward compatible with standard stereo but allowing to recover the individual tracks present in the mix can be of interest.

3 The Mixing Models

We present here the underlying model of all the methods we will consider in Sect. 4, as well as some generalizations.

We assume that the audio objects signals (or sources) are defined as M regularly sampled times series s_m of same length N. An audio object is thus understood in the following as a mono signal. Furthermore, we suppose that a mixing process produces a K-channel mixture $\{y_k\}_{k=1,\ldots,K}$ from the audio objects.

3.1 Linear Instantaneous Model

We first consider linear and time-invariant mixing systems. Formally, we suppose that each audio object s_m is mixed into each channel k through the use of some mixing coefficient a_{km}, thus:

$$y_k(t) = \sum_{m=1}^{M} y_{km}(t) \tag{1}$$

where

$$y_{km} = a_{km} \cdot s_m, \tag{2}$$

$\{y_{km}\}_{k=1,\ldots,K}$ being the (multi-channel) spatial image of the (mono) audio object s_m. In the stereo case where $K = 2$, we call this mono-to-stereo mixing.

We suppose that the mixing filters are all constant over time, thus leading to a time-invariant mixing system. We say that the mixing is linear instantaneous.

3.2 Convolutive Case

If the mixing coefficients a_{km} are replaced by filters, and the product in Eq. (2) is replaced by the convolution, we say that the mixing is convolutive. We can easily handle this case (see [12]) with the Short-Time Fourier Transform (STFT) representation if the length of the mixing filters is sufficiently short compared to the window length of the STFT, as:

$$Y_{km}(t, \omega) \approx A_{km}(\omega) S_m(t, \omega) \tag{3}$$

where $A_{km}(\omega)$ is understood as the frequency response of filter a_{km} at frequency ω. When the mixing process is linear instantaneous and time invariant, A_{km} is constant and the $K \times M$ matrix A is called the mixing matrix. When it is convolutive, this mixing matrix $A(\omega)$ is a function of ω. The mixing model can hence be written in the STFT representation as:

$$Y(t, \omega) \approx A(\omega) S(t, \omega) \tag{4}$$

where $Y = [Y_1, \ldots, Y_K]^\top$ and $S = [S_1, \ldots, S_M]^\top$ are column vectors respectively gathering all mixtures and sources at the time-frequency (TF) point (t, ω).

3.3 Non-linear Case

Of course, in real musical productions, non-linear effects such as dynamics compression are present in the mixing process. It has been shown in [24] that it is possible to

revert to the previous—linear—case by "moving" all the effects before the sum operation of the mixing model (more precisely, the effects placed in the processing chain after the sum operation merging the channels can be modified in order to be placed before this sum, see [24]). The problem with this approach is that it might lead to "altered" sound objects—i.e. "contaminated" by the effects—and thus harder to use for some active listening scenarios without noticeable artifacts. However, the methods presented in the next section have proved to be quite resistant to non-linearities of the mixing process.

3.4 Image-Based Model

In real-world conditions, the mixing process may be much harder to model [24]. Take for instance the stereo sub-mix of a multi-channel captured drum set, or the stereo MS recording of a grand piano. Then the solution is to not consider audio objects anymore but rather directly their spatial images [5]. Source separation consists then in undoing the sum of Eq. (1), to recover the M separate images $\{y_{km}\}_{k=1,\ldots,K}$ from the mixture $\{y_k\}_{k=1,\ldots,K}$. Each image has then the exact number of channels as the mix ($K = 2$ for a stereo mix). Such model will be referred to as stereo-to-stereo mixing. In such case, audio objects are not separated, but the modification of the separated images can still allow a substantial amount of active listening scenarios, including remixing and generalized karaoke. Respatialization, however, can be more difficult.

4 Informed Separation Methods

The objective of informed source separation is hence to compute some additional information that allows to recover estimates of the sources given the mixture $\{y_k\}_{k=1,\ldots,K}$. Depending on the method, these sources can be either the audio objects s_m or their spatial images $\{y_{km}\}_{k=1,\ldots,K}$ ($K = 2$ for stereo).

For the computation of the additional information, we assume that s_m and A are all available at the coder stage. Of course, the main challenge is to develop techniques that produce good estimates with an additional information significantly smaller than the one needed to directly transmit s_m.

Over the years of the project, we proposed several informed source separation methods. More precisely, this section presents the similarities, differences, strengths, and weaknesses of four of them. A detailed technical description or comparison is out of the scope of this chapter. The detailed descriptions of the methods can rather be found in [4, 14, 21, 23], while their comparison is done in [12].

4.1 Time-Frequency Decomposition

All the methods we propose are based on some time-frequency (TF) decomposition, either the MDCT or the STFT, the former providing critical sampling and the latter being preferred for the mixing model (see Sect. 3) and for filtering thanks to the convolution theorem.

Then, for each TF point, we determine the contribution of each source using several approaches and some additional information.

4.2 Additional Information

In the following, we assume that the encoder is provided with the knowledge of the mixing matrix A. However, this matrix may be estimated as demonstrated in [14]. This information may be used either directly or by deriving the spatial distribution of the sources. Then, our different methods have specific requirements in terms of additional information.

4.2.1 Source Indices

The first information used was the indices of the two most prominent sources, that is the two sources with the highest energy at the considered TF point. As explained below, this information can be used to solve the interference of the sources at this point. This information can efficiently be coded with $\lceil \log(M(M-1)/2) \rceil$ bits per TF point.

4.2.2 Source Energies

The information about the power spectrum of each source turned out to be extremely useful and more general. Indeed, if we know the power of all the sources, we can determine the two predominant sources. We can also derive activity patterns for all the sources. This information can efficiently be coded using for example the Equivalent Rectangular Bandwidth (ERB) and decibel (dB) scales, closer to the perception, together with entropy coding [4], or alternatively with Non-negative Tensor Factorization (NTF) techniques, as demonstrated in [13, 14].

4.3 ISS Methodologies

The majority of our ISS methods aims at extracting the contribution of each source from each TF point of the mix, at least in terms of magnitude, and of phase too for most of the methods.

The first method performs a local inversion [21] of the mix for each TF point, using the information of the two predominant sources in this point (as well as the knowledge of the mixing matrix). More precisely, at each TF point two sources can be reconstructed from the two (stereo) channels, by a local two-by-two inversion of the mixing matrix. This way, we get estimates of the magnitude and phase of the prominent sources. As discussed below, this method gives the best results with the Signal-to-Distortion Ratio (SDR) objective measure of BSSEval [25]. But the problem is that the remaining $M - 2$ sources exhibit a spectral hole (no estimated signal), which is perceived as quite annoying in subjective listening tests [4]. Also, this method requires the mixing matrix A to be of rank M.

The second method performs Minimum Mean-Square Error (MMSE) filtering [14] using Wiener filters driven by the information about the power of the sources (as well as the mixing matrix), the corresponding spectrograms being transmitted using either NTF or image compression techniques. Although this method produces results with a lower SDR, the perceived quality is higher, which matters to the listener. In contrast to the local inversion method, MMSE does not constrain as much the mixing matrix A and is therefore more flexible towards the mixing configurations. The separation quality, however, is much better when A is of rank M.

The third method performs linearly constrained spatial filtering [4] using a Power-Constraining Minimum-Variance (PCMV) beamformer, also driven by the information about the power of the sources (and their spatial distribution) and ensuring that the output of the beamformer matches the power of the sources (additional information transmitted in ERB/dB scales). In the stereo case ($K = 2$), if only two predominant sources are detected, the beamformer is steered such that one signal component is preserved while the other is canceled out. Applying this principle for both signal components results in inverting the mixing matrix (first method). Moreover, dropping the power constraint will turn the PCMV beamformer into an MMSE beamformer (second method). Otherwise, the PCMV beamformer takes advantage of the spatial distribution of the sources to produce best estimates than the early MMSE approach, at least with the PEMO-Q [8] measure, closer to the perception.

The fourth method performs iterative phase reconstruction and is called IRISS (Iterative Reconstruction for Informed Source Separation) [23]. It also uses the magnitude of the sources (transmitted in ERB/dB scales) as well as a binary activity map as an additional information to the mix. The main point of the method is to constrain the iterative reconstruction of all the sources so that Eq. (3) is satisfied at each iteration very much like the Multiple Input Spectrogram Inversion (MISI) method [7].

MISI estimates the phase of a set of sources composing a monophonic mixture using the amplitude of the STFT as an initialization. The method computes the so

called "re-mix" error between the original mix and the remix obtained using Eq. (3), this error is then redistributed on the phase of each steam and their amplitude is constrained to the prior known value.

Contrary to MISI, both amplitude and phase of the STFT are reconstructed in IRISS, therefore the remix error should be carefully distributed. In order to do such a distribution, an activity mask derived from the Wiener filters is used. The sources are reconstructed at the decoder with an initialization conditioned at the coding stage. It is noticeable that this technique is specifically designed for mono mixtures ($K = 1$), where it gives the best results, and does not yet benefit from the case $K > 1$.

The main remaining issue with the aforementioned methods is that their performance is bounded. Other methods [13, 19] are based on source coding principles in the posterior distribution of the sources given the mixtures and should permit to reach arbitrary quality provided that the bitrate of the additional information is sufficient.

4.4 Performance Evaluation

The quality performance of the system now reaches the needs of many real-life applications (e.g. industrial prototypes, see Sect. 5 below) with ongoing technology transfers and patents (for each of the four methods described above). The comparison of the current implementation of our four methods can be found in [12], for the linear instantaneous and convolutive cases (see Sect. 3), using either the objective SDR criterion of BSSEval [25] or the PEMO-Q measure [8], closer to perception. It turns out that the first method (local inversion) exhibits the best SDR (objective) results, while the third method (constrained spatial filtering) exhibits the best PEMO-Q (more subjective) scores; this was also verified in a formal listening test [4]. It is important to note that the complexity of these methods is low, enabling active listening in real time. Moreover, as shown in [12], the typical bitrates for the additional information are approximately 5–10 kbps for each audio object, which is quite reasonable.

5 Prototypes

Multiple versions of the DReaM system allow applications to uncompressed (PCM) and compressed (AAC/MP3/OGG) mixdown with mono-to-mono, mono-to-stereo, and stereo-to-stereo mixtures including artistic effects on the stereo mix [24]. However, the two main software prototypes (see below) are dealing with an uncompressed mix.

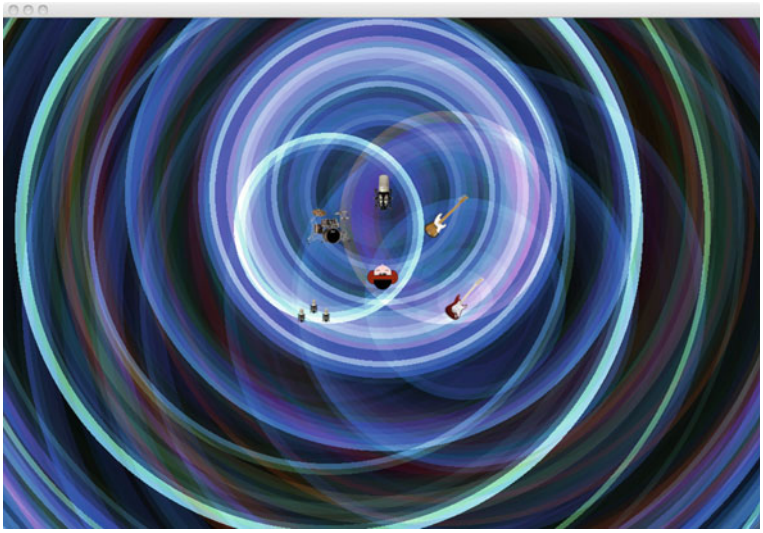


Fig. 3 From the stereo mix, the DReaM-RetroSpat player permits the listener (center) to manipulate 5 sources in the acoustic space (and to visualize the sound propagation)

5.1 DReaM-RetroSpat

We have presented in [15] a real-time system for musical interaction from stereo files, fully backward-compatible with standard audio CDs (see Fig. 3). This system manages the mono-to-stereo case and consists of a source separator based on the first DReaM method of Sect. 4 (local inversion) and a spatializer, RetroSpat [17], based on a simplified model of the Head-Related Transfer Functions (HRTF), generalized to any multi-loudspeaker configuration using a transaural technique for the best pair of loudspeakers for each sound source. Although this quite simple technique does not compete with the 3D accuracy of Ambisonics or holophony (Wave Field Synthesis—WFS), it is very flexible (no specific loudspeaker configuration) and suitable for a large audience (no hot-spot effect) with sufficient quality.

The resulting software system is able to separate 5-source stereo mixtures (read from audio CDs or 16-bit PCM files) in real time and it enables the user to remix the piece of music during playback with basic functions such as volume and spatialization control. The system has been demonstrated in several countries with excellent feedback from the users/listeners, with a clear potential in terms of musical creativity, pedagogy, and entertainment.

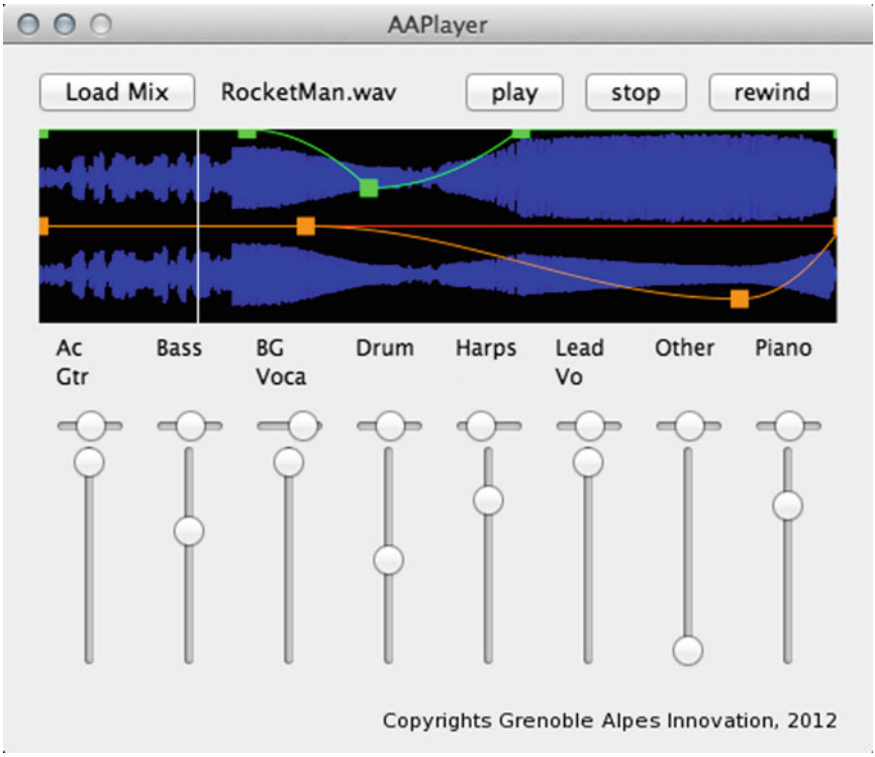


Fig. 4 Manipulation of a 8-source mix by the DReaM-AudioActivity player

5.2 DReaM-AudioActivity

The DReaM-AudioActivity prototype (see Fig. 4) targets consumer/prosumer applications of the ISS technologies issued of DReaM. The software is written in such as way that each separation method can be included as a separate C++ subclass, but only the MMSE filter method was implemented. This work was supported by GRAVIT (Grenoble Alpes Innovation) in collaboration with the DReaM team.

This prototype addresses the issue of studio music production, that is the stereo-to-stereo case. In some cases, the mix may not even be the exact sum of the stereo sources: dynamics processing can be applied and estimated a posteriori [24]. The coder performs, in almost real time, high-capacity watermarking of the separation information from the separated stereo tracks into the artistic mix coded in 16-bit PCM. The decoder performs offline reading of this watermark and performs the separation and re-mixing in real time. The number of tracks that can be included in the mix is only limited by the capacity of the watermark. Vector optimization of the audio processing core gives very low CPU usage during live separation and remixing. The end-user can then modify the volume and stereo panning of each source in real

time during playback. Automation of global and per track volume and panoramic is also possible. As always, the coded mix is backward compatible with standard 16-bit PCM playback software programs with little to no audio quality impact.

6 Conclusion

Originally thought as a way to interact with the music signal through its real-time decomposition/manipulation/recomposition, in the DReaM project the emphasis has been laid on the mixing stage, leading to source separation/unmixing techniques using additional information to improve the quality of the results. DReaM can also be regarded as a multi-track coding system based on source separation. Some of the techniques have been implemented in software prototypes, for demonstration purposes. These prototypes enable the user to perform, for instance, generalized karaoke and respatialization.

The initial aim was to give freedom to the listener, in the context of music industry, but artistic as well as industrial problems arose. For example, the artwork is sacred—it shall not be "altered". Also, the method requires studios recordings—involving copyright issues with studios/producers/majors. Finally, the method requires mastering the whole production chain—meaning entering Digital Audio Workstations (DAWs), which can hardly be done.

But DReaM has shown the possibility to produce a mix allowing source separation, backward compatible with legacy stereo, thus without the need of some multi-track format. We now plan to develop a patent-free version of the system, that could be used e.g. for storing/spreading multi-track sound archives within the standard stereo format.

Acknowledgements This research was partly supported by the French ANR *(Agence Nationale de la Recherche)*, within the scope of the DReaM project (ANR-09-CORD-006). "You may say I'm a dreamer, but am not the only one." (John Lennon—Imagine). Thus, the author would like to thank all the members of the project consortium for having made the DReaM come true.

References

1. Comon P, Jutten C (eds) (2010) Handbook of blind source separation—independent component analysis and applications. Academic Press
2. Fourer D, Marchand S (2013) Informed spectral analysis: audio signal parameter estimation using side information. EURASIP J Appl Signal Process 2013(1):178
3. Girin L, Pinel J (2011) Informed audio source separation from compressed linear stereo mixtures. In: Proceedings of the 42nd AES conference, Ilmenau, Germany, July 2011
4. Gorlow S, Marchand S (2013) Informed audio source separation using linearly constrained spatial filters. IEEE Trans Audio Speech Lang Process 21(1):3–13

5. Gorlow S, Marchand S (2013) Informed separation of spatial images of stereo music recordings using low-order statistics. In: Proceedings of the IEEE workshop on machine learning for signal processing (MLSP), Southampton, United Kingdom, September 2013
6. Gorlow S, Marchand S (2013) On the informed source separation approach for interactive remixing in stereo. In: Proceedings of the 134th AES convention, Roma, Italy, May 2013
7. Gunawan D, Sen D (2010) Iterative phase estimation for the synthesis of separated sources from single-channel mixtures. IEEE Signal Process Lett 17(5):421–424
8. Huber R, Kollmeier B (2006) PEMO-Q—a new method for objective audio quality assessment using a model of auditory perception. IEEE Trans Audio Speech Lang Process 14(6):1902–1911
9. ISO/IEC 23000-12 (2010) Information technology—multimedia application format (MPEG-A)—Part 12: Interactive music application format (IMAF)
10. Knuth KH (2005) Informed source separation: a Bayesian tutorial. In: Proceedings of the European signal processing conference (EUSIPCO), Antalya, Turkey, September 2005
11. Lepain P (1998) Recherche et applications en informatique musicale, chapter Écoute interactive des documents musicaux numériques, pp 209–226, Hermes, Paris, France, 1998 (in French)
12. Liutkus A, Gorlow S, Sturmel N, Zhang S, Girin L, Badeau R, Daudet L, Marchand S, Richard G (2012) Informed audio source separation: a comparative study. In: Proceedings of the European signal processing conference (EUSIPCO), Bucharest, Romania, August 2012
13. Liutkus A, Ozerov A, Badeau R, Richard G (2012) Spatial coding-based informed source separation. In: Proceedings of the European signal processing conference (EUSIPCO), Bucharest, Romania, August 2012
14. Liutkus A, Pinel J, Badeau R, Girin L, Richard G (2012) Informed source separation through spectrogram coding and data embedding. Signal Process 92(8):1937–1949
15. Marchand S, Mansencal B, Girin L (2011) Interactive music with active audio CDs. Lect Notes Comput Sci Explor Music Contents 6684:31–50
16. Marchand S, Badeau R, Baras C, Daudet L, Fourer D, Girin L, Gorlow S, Liutkus A, Pinel J, Richard G, Sturmel N, Zang S (2012) DReaM: a novel system for joint source separation and multi-track coding. In: Proceedings of the 133rd AES convention, San Francisco, California, USA, October 2012
17. Mouba J, Marchand S, Mansencal B, Rivet J-M (2008) RetroSpat: a perception-based system for semi-automatic diffusion of acousmatic music. In: Proceedings of the sound and music computing (SMC) conference, pp 33–40, Berlin, Germany, July/August 2008
18. Ozerov A, Févotte C (2010) Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation. IEEE Trans Audio Speech Lang Process 18(3):550–563
19. Ozerov A, Liutkus A, Badeau R, Richard G (2011) Informed source separation: source coding meets source separation. In: Proceedings of the IEEE workshop on applications of signal processing to audio and acoustics (WASPAA), pp 257–260, New Paltz, New York, USA, October 2011
20. Pachet F, Delerue O (1998) A constraint-based temporal music spatializer. In: Proceedings of the ACM multimedia conference, Brighton, United Kingdom
21. Parvaix M, Girin L (2011) Informed source separation of linear instantaneous under-determined audio mixtures by source index embedding. IEEE Trans Audio Speech Lang Process 19(6):1721–1733
22. Pinel J, Girin L, Baras C, Parvaix M (2010) A high-capacity watermarking technique for audio signals based on MDCT-domain quantization. In: Proceedings of the international congress on acoustics (ICA), Sydney, Australia, August 2010
23. Sturmel N, Daudet L (2013) Informed source separation using iterative reconstruction. IEEE Trans Audio Speech Lang Process 21(1):178–185
24. Sturmel N, Liutkus A, Pinel J, Girin L, Marchand S, Richard G, Badeau R, Daudet L (2012) Linear mixing models for active listening of music productions in realistic studio conditions. In: Proceedings of the 132nd AES convention, Budapest, Hungary, April 2012
25. Vincent E, Gribonval R, Févotte C (2006) Performance measurement in blind audio source separation. IEEE Trans Audio Speech Lang Process 14(4):1462–1469

Finding Music in Music Data: A Summary of the DaCaRyH Project

Oded Ben-Tal, Bob L. Sturm, Elio Quinton, Josephine Simonnot and Aurelie Helmlinger

Abstract The international research project, "Data science for the study of calypso-rhythm through history" (*DaCaRyH*), involved a collaboration between ethnomusicologists, computer scientists, and a composer. The primary aim of *DaCaRyH* was to explore how ethnomusicology could inform data science, and vice versa. Its secondary aim focused on creative applications of the results. This article summarises the results of the project, and more broadly discusses the benefits and challenges in such interdisciplinary research. It concludes with suggestions for reducing the barriers to similar work.

1 Introduction to the DaCaRyH Project

The *DaCaRyH* project ("Data science for the study of calypso-rhythm through history") was a 22-month long international collaboration of ethnomusicologists (Helmlinger, Spielmann), music signal processing engineers (Sturm, Quinton), music

O. Ben-Tal
Kingston University, London, UK
e-mail: o.ben-tal@kingston.ac.uk

B. L. Sturm (✉)
Royal Institute of Technology KTH, Stockholm, Sweden
e-mail: b.sturm@qmul.ac.uk

E. Quinton
Universal Music Group, Santa Monica, USA
e-mail: elio.quinton@umusic.com

J. Simonnot · A. Helmlinger
CREM-LESC, UMR7186, CNRS, Nanterre, France
e-mail: Josephine.SIMONNOT@cnrs.fr

A. Helmlinger
e-mail: aurelie.helmlinger@cnrs.fr

R. Bader (ed.), *Computational Phonogram Archiving*, Current Research in Systematic Musicology 5, https://doi.org/10.1007/978-3-030-02695-0_9

archivists (Simonnot), software developers (Parisson), and a composer (Ben-Tal). The project was jointly funded by national organisations in both France (LABEX) and the UK (Arts and Humanities Research Council), and ran from February 2016 to November 2017. The overarching impetus of the project was to explore how ethnomusicology can inform the development of useful and meaningful computational methods in music recording archives. This is in contrast to past approaches appearing in the literature, where (ethno)musicologists are brought a variety of computational tools and asked to use them; or where computational methods are applied to music recording collections without any involvement of (ethno)musicologists [23].

DaCaRyH was motivated by a desire of its members to familiarize themselves with the emerging field of computational ethnomusicology, and to contribute to it. *DaCaRyH* had three principal objectives:

1. *Ethnomusicological*: To enrich the domain of ethnomusicology by integrating data science and music information retrieval (MIR) methods [20] into ethnomusicological archives and research practices[1];
2. *Computational*: To enrich data science and MIR research by integrating real ethnomusicological use cases and requirements into the design of intelligent systems;
3. *Creative*: To study the concept of musical style through a comparative diachronic analysis of a music dataset, and then the creative transformation of features extracted from the same dataset into an invented, imaginary style.[2]

The particular music which served as the starting point of the project was the music tradition of the steelband calypso,[3] and steelband music as carnival arts [2, 4, 9, 14] and in particular, the use of rhythm across time over a period of about 50 years in the annual *Panorama competition*. This competition was created shortly after Trinidad and Tobago became independent in August 1962 as part of a policy governmental promoting masquerade, calypso traditional narrative ballads. The state-sponsored Panorama competition has since become the most prestigious steelband stage performance in Trinidad. Owing to both its prominent and prestigious position, the Panorama competition has had a prime influence on the development of steelband music [5, 24].

In this article we review the results of *DaCaRyH*, and discuss lessons we learned in the process. Section 2 reviews the context of the project, and Sect. 3 discusses its groundwork. In Sect. 4, we describe results of this project in the computational ethnomusicology domain, and in Sect. 5, we summarise our creative work. We then draw some conclusions about such interdisciplinary and collaborative research.

[1]Music information retrieval encompasses computational methods for extracting, accessing and using information in collections music recordings. Examples include estimating the tempo of music in an audio recording, determining the key of a notated piece of music, and transcribing an audio music recording into notation.

[2]A "dataset" is a collection of data. In our case, it is a collection of audio music recordings.

[3]Calypso is a style of Caribbean music that originated in Trinidad and Tobago. For a more detailed description of Calypso, see for example [3].

2 Context of the Project

DaCaRyH arises from a somewhat contentious debate regarding "data science", "big data" and the computational study of culture [7, 13, 15, 16, 21, 22, 25, 26, 30]. Mauch et al. [16] argues strongly for the utility of "big data tools" for the scientific study of music. In their work, they apply computational tools to analyse excerpts of American popular music and discover three "stylistic revolutions" between 1960 and 2010. This follows on the heels of other "big data" music studies. Serra et al. [22] investigated changes along three dimensions of Western popular music over 50 years and discover in their results that the "pitch sequences" have become more restricted, the "timbral palette" has become less diverse, and that the "average loudness" has increased. Another study [21] analyses the statistics of tempo, mode, duration, artist gender, and recording year of Top-40 songs since 1960, and finds tempo to be decreasing along with the use of major mode. More recently, Weiß et al. [31] explored the stylistic evolution of Western classical music from 1660 to 1975 by comparing tonal descriptors such as intervalic content, chord transitions, and "tonal complexity". Their findings essentially confirm broad observations of musicologists with regards to stylistic development and boundaries during the period of study, which suggests the underlying computational framework is musically meaningful.

Some of these findings produced using "data science" and "big data", however, are not broadly persuasive. Though they arise from computational studies of a vast number of music recordings that would strain any human listener, the "objectivity" and "relevance" of the methods employed have been challenged. A talk at a national meeting of the American Musicological Society [30] and an editorial [7]—both by musicologists—bemoans the "science" of such studies, the aggrandisement of such findings by the media, and the lack of participation in such studies of musicologists, i.e., those who actually *study* music and culture. Also problematic are implications that what musicologists do is not a "scientific study" of music, in contrast with "big data" approaches, which themselves have inherent biases. Recent surveys [25, 26] show an enormous amount of work has been produced by music data scientists but that much of it treats music superficially, and produces conclusions with questionable scientific validity. The presentation by Marsden [15] at the 2015 workshop *Music Similarity: Concepts, Cognition and Computation* also criticises the treatment of music as documents from which to extract a diverse number of statistics of unknown relevance, instead of treating music as artefacts or records of human culture with rich contexts.

From the published literature, it seems that musicologists have so far rarely been invited to take a seat at this "big data buffet", but this is beginning to change. The AHRC-funded project "Transforming Musicology"[4] piloted ways of using digital tools in musicological studies. One project it supported was the "Digital Music Lab" (AH/L01016X/1),[5] which sought to build tools to facilitate "big data musicology" [32]. A case study of applying "big data" to musicology is given by the project "A Big

[4] http://www.transforming-musicology.org.

[5] http://dml.city.ac.uk.

Data History of Music" (AH/L010046/1).[6] The US musicologist Huron [13] argues that it is a "moral imperative" to welcome "big music data," and for new trainees to become versed in the methods of data science and statistics. This call to action is also echoed by [30].

This context motivated the approach we proposed in *DaCaRyH*: instead of starting by asking what (ethno)musicological questions might be answered by specific tools and techniques of music data science, we begin from the ethnomusicological questions, and then attempt to create bespoke tools that could help address them. The resulting learnings may then feedback and inform MIR practices as well as feeding our creative strand.

3 Groundwork

3.1 Steelband's Calypso: Panorama Recordings

Calypso is a polysemic term referring to a group of related styles originating from Trinidad and Tobago. In its most common sense, it is a song of social commentary, with a specific rhythmic accompaniment, and performed in the western tonal system. The calypso beat is also the polyrhythmic accompaniment, that can be played independently from the song. Historical researches show the earlier forms were responsorial, played with a polyrhythmic accompaniment, in the carnival procession [3, 18]. The 19th century musical form called kalinda first played both indoor (in tents) and in the streets, evolved differently according to the contexts, that oriented organological and musical changes, within a certain stylistic homogeneity. The calypso song appeared in the late 19th century, and is the heir of the indoor style, that adopted Western instruments, and the street songs and polyrhythms later evolved to steelbands, after several organological changes: from drum to tamboo bamboo (stamping tubes polyrhythm), and from bamboos to metallic second hand instruments.

The adoption of metal allowed a dramatic musical innovation in the 1940s: the players started to differentiate several pitches on the same surface [5, 24]. This inclusion of pitches in the polyrhythm launched the era of "steelpans" (or "pans") melodic idiophones made of the 55-gallon oil drums and main instrument of steelbands. A music style had led to the invention of a new musical instrument family, that became popular in the US under the name "steeldrums". The making of steelpans gradually professionalized, and is now highly specialized: to make a pan, the maker (called tuner) sinks the top of the drum, shapes the notes, cuts the body of the drum. Then the instrument is burnt, tuned with a hammer, sometime chromed. Tuners perform nowadays generally the "harmonic tuning", the tuning of some overtones following the harmonic series, which had a great impact on the tone of the steelpans [11]. Although the earlier pans were made of one drum (and they still do in the "single

[6] http://gtr.rcuk.ac.uk/project/3510829B-EAE9-48DC-A723-8093D92CAD60.

pan bands"), the number of drums per musician can—in the fully chromatic "conventional bands"—reach up to twelve for the lowest range instrument.

Amazingly, although calypso music became a natural music style for steelbands, it not the only one: steelbands have rapidly played in various contexts, in addition to the carnival procession. They started soon to perform a large variety of tunes, from western classical music to pop and jazz [24]. But a political event, the Independence of Trinidad and Tobago (1962), happened to influence the music scene in the steelband movement: a new competition, the Panorama, was created by the authorities as soon as 1963. Scheduled right before the carnival, its rules constrain the band to play a calypso. Panorama quickly became the major event amongst steelbands, gathering hundreds of bands (all categories included), composed of up to a hundred players each. Interestingly, the competitive context happened to shape a particular style of music, belonging to the calypso family: the Panorama tune. The rules constraint the bands to perform the arrangement of a calypso. The arranger task is to select a song, and arrange it into the Panorama structure, which is influenced by the sonata form. An introduction is composed, then the verse and chorus of the calypso are presented, followed by various variations including modulations, a change into minor mode, a sort of climax called a jam, another exposition to the theme, and a coda.

Steelband music offers the rare possibility to observe the evolution of a music style—including a new family of instrument and their acoustical characteristic—almost from its creation. The Panorama competition is a consistent and very meaningful category in Trinidad and Tobago, where the steelpan is a true emblem: it was officially declared the National Instrument in 1992 [5, 18].

3.2 Session Music Transcriptions

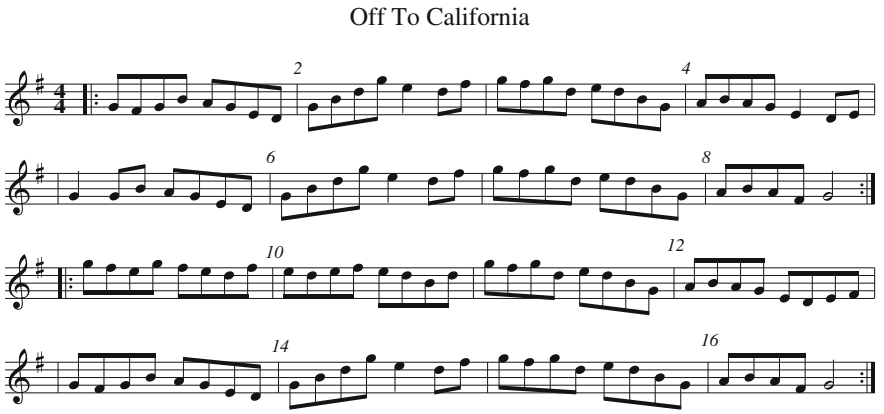
The creative strand of the project envisioned the transformation of features extracted from the recorded steelband music dataset to create a form of music different from calypso but informed by it. Fairly soon it became apparent that extracting useful features from the noisy recordings will be very difficult. Rather then delay the work on this aspect we decided to apply the same concept to another dataset as a proof of concept: What creative paths open with access to a symbolic representation of a large collection of traditional music?

We settled on using transcriptions of traditional music referred to as "session" music—traditional music typically accompanying dance from the UK, Ireland and France, but also including other traditions, e.g., Cajun. The website thesession.org serves as a hub for an international community of enthusiasts for this kind of music, and provides thousands of such transcriptions. Registered users post transcriptions of tunes in a textual representation called "ABC notation".[7] An example of one transcription from this collection is below:

[7] http://abcnotation.com/wiki/abc:standard:v2.1.

```
T: Off To California
R: hornpipe
M: 4/4
K: Gmaj
|:GFGB AGED|GBdg e2df|gfgd edBG|ABAG E2DE|
|G2GB AGED|GBdg e2df|gfgd edBG|ABAF G4:|
|:gfeg fedf|edef edBd|gfgd edBG|ABAG EDEF|
|GFGB AGED|GBdg e2df|gfgd edBG|ABAF G4:|
```

The fields in this transcription denote the title (`T:`), dance (`R:`), meter (`M:`), mode (`K:`), and finally the notes and repeats in the tune. This kind of music is typically learned by ear, and so such a transcription mostly serves as a reminder (which is one of the original motivations that drove the development of ABC notation). The way a tune is played is contingent on the dance it accompanies and the instrument that is playing it, and may not be explicit in a transcription. For instance, the above transcription converted to staff notation appears as the following:



However, when this tune is played to accompany a hornpipe the quaver rhythm is broken (offbeat quavers are shortened) and beats 2 and 4 are emphasised.

This crowd-sourced data contains more then 23,000 transcriptions of melodies, many of which are different versions of the same tune. The tune above has seven different transcriptions in the dataset,[8] each with slight variations, e.g., ornamentations, an explicit hornpipe rhythm, first and second endings, and so on. The size of this dataset makes it amenable to modeling with neural networks, which we describe in Sect. 5.

[8] https://thesession.org/tunes/30.

3.3 CREM and Telemeta

The audio archives of the CNRS Musée de l'Homme gather commercial and field recordings of music and oral traditions from all around the world, from 1900 to the present. These archives are among the most important in Europe in terms of quality, quantity and diversity. The "entre de Recherche en Ethnomusicologie", a research team of the National Center for Scientific Research in France (LESC, UMR 7186), manages this intangible heritage.

Most of the ethnomusicological archives are sound recordings resulting in unique challenges in preserving and managing such a collection in spite of the technological evolution. The temporal nature of audio-visual materials used in these fields raises special issues. Because they are used for research, they must be both well indexed and also easily accessible. They must also be managed in such a way as to provide their associated metadata and controlled access. In addition, it is important to be able to visualize the sound in order to navigate through the files and to annotate them precisely, which is not as simple as annotating text or image documents. At the same time, the French academic institutions require a wide access to the raw data to justify the continued support for research projects and continuing digitisation efforts. So the challenge is to manage the intellectual property rights as well as the technical issues.

The Research Center for Ethnomusicology (CREM) and the Laboratory of Musical Acoustics (LAM), have been working together since 2007 to design an innovative and collaborative tool for easy indexing of sound files. This tool, called Telemeta, a multimedia Web platform based on an open-source software designed by the Parisson Company, has been online since 2011. Now, access to the audio archives of the CNRS—Musée de l'Homme is offered with a streaming player.[9] More than 43,000 recordings and metadata collected throughout the world since 1900 are available on line in the database for research and experimentation in ethnomusicology. Since 2013, the aim is to enrich Telemeta platform with MIR tools used by data scientists (from IRIT, LIMSI, LaBRI, France) in order to pave the way for a semantic search engine. To be effective and in phase with the new audio technologies, analysis tools are expected to improve musical research activities on the web database. The reflection collectively engaged by engineers and researchers on the use of the sound archives database led us to set up a national project called DIADEMS (Description, Indexation, Access to Ethnomusicological and Sound Documents, 2013–2016). These new tools are used by the CREM staff for indexing since 2015. Now, to index and to segment the audio content, we are helped in this long and careful process with on line tools, like start recorder detection, speech/singing detection, monophonic/polyphonic parts detection, etc. With those tools and temporal annotation markers, the CREM is developing a new collaborative work between the sound engineer and the archivist to identify the content the sound files in an efficient way.

[9] http://archives.crem-cnrs.fr.

4 Ethnomusicological Application

One of the work threads in DaCaRyH consisted in an interdisciplinary case study combining traditional and computational methodologies to study Trinidad steelband music in a collection of recordings of the annual Panorama competition spanning over 50 years [17]. A number of facts and trends have been identified regarding Trinidad steelband music in the ethnomusicology literature,[10] while some other hypotheses formulated have not been addressed with traditional methodologies. As a result, we investigated these ethnomusicological research questions through the computational lens of MIR methods [20] to facilitate the realisation of quantitative and labour intensive studies. More specifically, we focused on trends of tempo, tuning and dynamic range over a period of 50 years of the Panorama competition. We sought to study three research questions:

1. Does the tempo of winning Panorama performances tend to increase over time?
2. Do arrangers use increasingly large dynamic ranges?
3. Has there been a change in the sonic properties of steelbands?

We assembled a dataset of recordings of most of the steelband calypsos/socas[11] performed by the winners (first, second and third places) for each yearly Panorama competition between 1963 and 2015 (except for the year 1979, in which the competition was boycotted). Our dataset contains 93 recordings, each one being typically 8–10 min long, totaling about 14hrs of audio. All the digital recordings of this dataset were extracted from the "CNRS—Musée de l'Homme" sound archives,[12] which are accessible via the Telemeta platform [6],[13] as described in Sect. 3.3. This musical data includes field recordings as well as published recordings, which are a mixture of digitised analog recordings and digital recordings.

The details of our work are given in [17], but in summary we automatically extracted features, such as tempo, loudness or tuning frequency, from every audio recording in order to provide a quanitative measurment to shed some light on our research questions. Not all MIR descriptors extracted are equally reliable and some did not provide robust enough evidence to draw conclusions. But when these quantitative measurment were conclusive, our findings are in line with what appears in the ethnomusicological literature.

For instance, Fig. 1 shows the estimated tempo of each recording in our collection in each competition year. We noticed that some of the estimates are below 110 bpm or above 140 bpm—which are unusual tempi for this music. Manual inspection reveals that most of these come from errors of the automatic tempo estimation procedure. Nevertheless, in the vast majority of cases tempo automatically extracted matches well with human hearing and may therefore be considered as reliable. On the other

[10] See for example [1, 4, 11].

[11] In the dictionary of the English/creole of Trinidad and Tobago edited by Lise Winer soca is defined as "a type of calypso-based music, with a fast dance beat, and party lyrics.".

[12] http://archives.crem-cnrs.fr/.

[13] http://telemeta.org/.

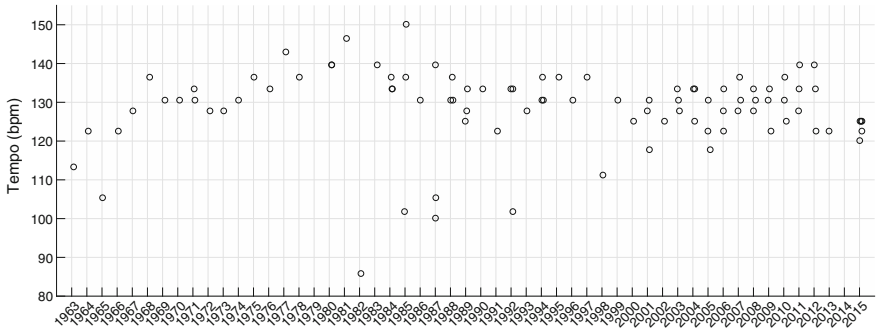


Fig. 1 From [17], the estimated tempo of each recording in our collection

hand, when it comes to dynamic range and tuning, the results obtained from MIR tools on such a noisy dataset are much less reliable, and therefore do not constitute enough evidence to support a conclusion regarding our related research questions. One reason for this unreliability comes from the recording conditions: these are live events many times recorded from positions closer to the audience than the musicians. We find in several recordings that the noise from the audience has more acoustic power than the music. While the human auditory system is remarkably effective at separating such sources, this is not the case with (current) MIR tools. As a result, the numbers calculated by the automatic estimation are not musically relevant.

Music recording corpora used for ethnomusicological research typically are either ethnographic recordings or field recordings (both in our case), which are produced in very heterogeneous conditions that are mostly uncontrolled. It is often not possible to evaluate the impact of these recording conditions on the properties of the musical recordings. Our study is a good demonstration of how some MIR tools are more impacted than others by the quality of audio. Some estimates produced by MIR tools on such corpora may only have limited reliability, while some others may be reliable. As we have highlighted, results of computational analyses carried out on such a dataset should therefore be interpreted with due care when no means of assessing the reliability of feature estimation are available. In addition, our study is also a demonstration of the fact that such datasets provide a challenge for the design of MIR tools robust against uncontrolled recording conditions.

5 Creative Application

In [27, 28], we describe our application of machine learning to the large collection of folk melodies described in Sect. 3.2. We name this software “folk-rnn”. The method we used—recurrent neural networks (RNN, [12])—takes the existing context to predict the next step in a sequence. Our first version applies this directly to the characters of the ABCnotation, while the second version is based on tokens. To illustrate

the difference consider the common-time meter representation: `M:4/4`. In the first version it will appear as a sequence of five characters: "`M`", "`:`", "`4`", "`/`", and "`4`". In the second version it is treated as a single symbol (thus "token"). Similarly a bar repeat sign "`:|`" would be two characters but only one token. We can see that from these two examples that the "`:`" character means different things in ABC notation. While an RNN might learn to take into account such differences in context, a token-based representation brings the transcriptions closer to the music these sequences encode.

After training, the resulting model is able to produce new transcriptions that share many of the characteristics of the training set. As our evaluation of this model shows [27] the model encodes some compositionally useful aspects of the style including correctly counting bars, basic phrase structure, repetition and variation of patterns, and cadence points (even though the cadences themselves are not consistently good). Furthermore, experienced performers who work within this tradition did not have difficulties locating good tunes within the large collection of produced transcriptions.[14] Musicians performed some of these tunes together with traditional tunes in several concerts and workshops we organised as well as at a pub session in London. One of the musicians commented that the model produced interesting and stylistically appropriate patterns that he did not come across before. One interpretation of this is that machine learning can augment this musical form—allowing musicians to explore new corners of the musical space of plausible session tunes. At the same time we must also acknowledge that some would consider this a distortion of the tradition.

We explored the creative potential of this approach beyond the domain of origin. We asked two musicians who do not perform folk music if they could find transcriptions they could play in a concert. Both John Hughes (double bass) and Torbjorn Hultmark (trumpet/soprano trombone) are experienced improvisors who regularly perform in different contexts: from orchestral works to Jazz and free improvisations. Both play instruments not typically included in session music. Yet both were able to curate and adapt transcriptions from our published collection and make them work.[15] This is foremost a testament to the musicality and creativity of performing musicians. But it also points to the co-creative potential of such Artificial "Intelligence" methods in the domain of music. To further explore this potential we created a web-based interface to our model. folkrnn.org allows users to generate transcriptions using our folk-rnn model. Users can generate and curate from the outputs or modify initiating parameters (such as meter, mode or opening notes) to explore the range of tunes the model is able to produce. Users will eventually be able to archive tunes they create with the tool using an online repository. There they will also be able to participate in forum discussions, browse and 'like' tunes other users have contributed and, hopefully, give and receive feedback about the tool and its potential for creating music. Our aim is to learn from the way others use the model

[14] https://highnoongmt.wordpress.com/2018/01/05/volumes-1-20-of-folk-rnn-v1-transcriptions/.

[15] Hultmark: https://www.youtube.com/watch?v=4kLxvJ-rXDs; Hughes: https://www.youtube.com/watch?v=GmwYtNgHW4g.

for creative purposes to further develop this as a composition assistance tool. We will be holding a composition competition—aimed primarily at students but open to everyone—with the winning piece performed in a concert in October 2018 in London.

Both Sturm and Ben-Tal used the folk-rnn model to compose new pieces [29]. Sturm based several compositions on selected system outputs which he then arranged electronically or acoustically. He also worked interactively with the system, seeding it with an initial sequence and then curating from the resulting outputs. Ben-Tal used the system to produce pre-compositional material which was substantially edited and adapted in the composition process itself. The generation of the material was also interactive—fine tuning parameters and seeding sequences. The aim was to shift the outputs away from the core style of the model to better fit Ben-Tal's compositional idiom. By doing that we discovered that our assumptions about what musical learning took place in the training phase were wrong. Starting the generation process from an initial sequence of tokens that deviated from the style of the training data, for example non-modal patterns, the continuation often did not contain the correct number of beats in a bar. The model was not able to apply repetition and variation on unfamiliar patterns resulting in "noodling", often musically unrelated to the initial material. In other words, the ability of the model to encode relevant features is highly circumscribed and not really musical. At least not musical in human terms. The creative research strand of our project, therefore, brought to light aspects that were hidden until we started prodding our model beyond a validation of the tool with statistics.

6 Lessons Learned

6.1 Differences in Research Practices Between Engineering and Ethnomusicology

Collaboration between MIR, which belongs to engineering and computer sciences, and ethnomusicology, related to both anthropology and musicology, highlights a set of differences in the scientific cultures and practices. These differences appear of course in the research methods, but also in the terminologies and practices: while MIR researchers follow engineering scientific standards which focus on benchmarks and statistical significance, ethnomusicologists collect information through "participant observation" resulting in rich, multifaceted data most of it irreducible to numbers. Cross-disciplinary communication is therefore crucial to bridge the gap and the understanding of the potential of each other approach, and to familiarize with each other terminology, concepts, and aims. This communication is made more complex by the differences in training between the disciplines, but made possible because of the common interest—the music.

Some of the scope is shared, but each discipline is driven by concerns that may lie outside the shared scope. For instance, in the case study discussed here, both ethnomusicology and MIR are interested in timbre, loudness and tempo. But in ethnomusicological literature, it is discussed as the effects of rivalry [5, 10]. The rivalry, a relational characteristic observed in Trinidad and Tobago steelbands, is the aspect ethnomusicologists have focused on: in an anthropological perspective, this is the feature that allows to understand both extra-musical behaviors and musical characteristics and evolution. Tempo, timbre and loudness interest ethnomusicologists for what these features express about musical culture: they are a matter of an additional interpretation. As Rouget puts it, "la musique, c'est toujours plus que la musique" [19], meaning "music is always more than just the music" and the ethnomusicological focus is wider than MIR. At the same time the scope of MIR is wider in its aims to provide answers about music in general, not just the music of a particular culture. Reconciling the differences in academic cultures requires learning to formulate (or re-formulate) research questions in ways that take into account the other field perspectives.

6.2 Difficulty in Forming Questions Compatible with Tools

Once we narrow the scope of both disciplines to the common ground, research questions fitting in this space have to be found. A pluri-disciplinary collaboration is materialized by different aspects: research questions, methods, bibliography. Methods and bibliographic work can be just added, so they naturally mix by addition. But research questions are less easily pluri-disciplinary: the orientation of the questions depends on one's scientific background. Goody [8] has shown how the characteristics of the media have a very deep impact on the content of it. In our case, the tools are not just a consequence of the research questions, they also influence them. The promise of computational ethnomusicology is that with new tools come new avenues for research. Without a background in the engineering sciences, one of the difficulties has been for the ethnomusicological part of the partnership to gain a sufficient understanding of the possibilities, as well as the limitations, offered by MIR tools to suggest ethnomusicological research questions fully adapted to the tools. Cross-readings, and extended exchanges are of course the way to overcome these hurdles: it is necessary for ethnomusicologists to be able to assimilate this new methodological arsenal in order to effectively formulate anthropological questions. Similarly, researchers working in the MIR domain would benefit from a better understanding of the wider context of (ethno)musicological knowledge to look beyond extracted features, data-sets, and ground truths.

6.3 Validation of Computational Tools

Computational tools designed for and applied to studying music have to be presented with a fair evaluation of their reliability or accuracy for the task in a *particular* context. In our case, manual validation shows that the tempo estimation tool is quite reliable. Most of the calculated values match what we identify from listening—the MIR tool echoes human interpretation and therefore can be considered reliable for the task. But manually double checking each MIR-produced result is little progress on estimating tempo manually to begin with. To make the tools really useful to non-MIR researchers they need to come with a "health warning". We need a way to estimate the fitness of the tool to the items it is unleashed on. Our work on estimating tuning frequency serves as an illustration of the problem. At first, calculations from MIR tools supported our hypothesis. Further investigations, informed by an understanding of the inner working of these methods, led us to conclude that these numbers are artefacts and are not reliable.[16] Either any computational ethnomusicology work will depend on a MIR "technician" or the tools themselves will need to present the user with more contextual information.

7 Conclusion

Our project was funded by a Franco-British initiative that recognises the importance of collaboration—across disciplines and cultures—to foster research. Our experience over the last two years is that the challenges with this kind of work are often more subtle and less visible than is commonly recognised. Disciplines shape not just terminology and methodology, they also influence career trajectories and pressures, academic cultures, and similar frames of reference. The two strands of work summarised in this paper illustrate two approaches to such collaborative work.

The work described in Sect. 4 fits into an interdisciplinary model where researchers in two domains bring their individual expertise to bear on a problem. The work is based on a limited dataset, where field researchers have empirical observations. Their expertise about this music allows them to posit hypothesis that can be addressed by MIR tools. At the same time, the expertise of the engineers is needed not just to apply the tools to the dataset but also to evaluate those results. Understanding the mechanisms of feature extraction is needed in order to understand how those may interact with the items in the collection.

Our work with machine learning and folk music (Sect. 5) is an example where the disciplines are much more intertwined. This outcome is the fruit of a collaboration that started almost two years before the beginning of the DaCaRyH project. Extensive discussions, most of them not tied to an immediate goal (write a paper; apply for a grant), enabled us to discover where our shared interests are and where our expertise is fruitfully complementary. The longer duration also allowed us to understand enough

[16] See further details in [17].

of each other's methods and concerns to be able to better translate our ideas between the different academic cultures.

That suggests taking a more patient approach to allow interdisciplinary collaboration to flower. This applies to the researchers who embark on such work, bodies who fund such research, as well as academic institutions which set expectations for research "productivity". At the same time we should also expand the training in research methods in our Ph.D. programmes. Computational methods, including both MIR and machine learning, have a role to play in future research in music. Unless we who research music and know it inside-out engage with those tools others will. We need to incorporate MIR tutorials into (ethno)musicology training programmes. We need to understand the potential and limitation of these as research tools. We need to showcase instances were computation serves musical purposes and highlight cases where research is getting music wrong.

At the same time, the MIR research community should consider more carefully the implicit assumptions that underlie the mechanisms they develop and include those when releasing tools. Simultaneously, MIR researchers should engage more with musicologists, ethnomusicologists, and music theorists. While it is clear that research in MIR will continue to be shaped primarily by the demands of commercial music, music departments hold stores of knowledge that the pertinent to the concerns of MIR research, yet remain largely untapped.

Acknowledgements Florabelle Spielmann, Ghislaine Glasson Deschaumes, Andrew Thompson.

References

1. Aho WR (1987) Steel band music in Trinidad and Tobago: the creation of a people's music. Lat Am Music Rev/Revista de Msica Latinoamericana 8(1):26–58
2. Birth KK (2008) Bacchanalian sentiments; musical experiences and political counterpoints in Trinidad. Duke University Press
3. Cowley J (1998) Carnival, canboulay and calypso: traditions in the making. Cambridge University Press
4. Dudley S (2002) The steelband "Own Tune": nationalism, festivity, and musical strategies in Trinidad's panorama competition. Black Music Res J:13–36
5. Dudley S (2008) Music from behind the bridge: steelband spirit and politics in Trinidad and Tobago (illustrated edition). Oxford University Press Inc
6. Fillon T, Pellerin G, Brossier P, Simonnot J, La Dfense N (2014) An open web audio platform for ethnomusicological sound archives management and automatic analysis. In: Workshop on folk music analysis (FMA2014), p 36
7. Fink R (2013) Big (bad) data (editorial). Musicology now (online)
8. Goody J (1977) The domestication of the savage mind. Cambridge University Press
9. Guilbault J (2007) Governing sound: the cultural politics of Trinidad's carnival musics. University of Chicago Press
10. Helmlinger A (2011) La virtuosité comme arme de guerre psychologique. Ateliers d'anthropologie 35
11. Helmlinger A (2012) Pan jumbie. Socit d'ethnologie, Mmoire sociale et musicale dans les steelbands (Trinidad et Tobago)
12. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780

13. Huron D (2013) On the virtuous and the vexatious in an age of big data. Music Percept 31(1):4–9
14. van Koningsbruggen PH (1997) Trinidad carnival: a quest of national identity. Caribbean
15. Marsden A (2015) Music similarity. In: Presentation at music similarity: concepts, cognition and computation. http://www.lorentzcenter.nl/lc/web/2015/669/presentations/Marsden.pptx
16. Mauch M, MacCallum RM, Levy M, Leroi AM (2015) The evolution of popular music: USA 1960–2010. R Soc Open Sci 2(5). https://doi.org/10.1098/rsos.150081
17. Quinton E, Spielmann F, Sturm BL (2017) Computational ethnomusicology for exploring trends in Trinidad steelband music through history. In: Proceedings of CMMR
18. Regis L (1999) The political calypso: true opposition in Trinidad and Tobago, 1962–1987. University Press of Florida
19. Rouget G (1995) Ethnomusicologie d'un rituel. la représentation, ou de velasquez à francis bacon. L'Homme 35(133):77–97
20. Schedl M, Gomez E, Urbano J (2014) Music information retrieval: recent developments and applications. Found Trends Inf Retr 8(2–3):127–261
21. Schellenberg EG, von Scheve C (2012) Emotional cues in american popular music: five decades of the top 40. Psychol Aesthet Creat Arts 6(3):196–203
22. Serra J, Corral A, Boguna M, Haro M, Arcos JL (2012) Measuring the evolution of contemporary western popular music. Sci Rep 2. https://doi.org/10.1038/srep00521
23. Spielmann F, Helmlinger A, Simonnot J, Fillon T, Pellerin G, Sturm BL, Ben-Tal O, Quinton E (2017) Zoom arrière: L'ethnomusicologie à l'ère du Big Data. Cahiers d'ethnomusicologie 30:9–28
24. Stuempfle S (1995) The steelband movement: the forging of a national art in Trinidad and Tobago. University of Pennsylvania Press
25. Sturm BL (2014a) The state of the art ten years after a state of the art: future research in music information retrieval. J New Music Res 43(2):147–172
26. Sturm BL (2014b) A survey of evaluation in music genre recognition. In: Nürnberger A, Stober S, Larsen B, Detyniecki M (eds) Adaptive multimedia retrieval: semantics, context, and adaptation, vol 8382, pp 29–66. LNCS
27. Sturm BL, Ben-Tal O (2017) Taking the models back to music practice: evaluating generative transcription models built using deep learning. J Creat Music Syst 2(1)
28. Sturm BL, Santos JF, Ben-Tal O, Korshunova I (2016) Music transcription modelling and composition using deep learning. In: Proceedings Conference Computer Simulation of Musical Creativity, Huddersfield, UK
29. Sturm BL, Ben-Tal O, Monaghan Ú, Collins N, Herremans D, Chew E, Hadjeres G, Deruty E, Pachet F (2018) Machine learning research that matters for music creation: a case study. J New Music Res (in press). https://doi.org/10.1080/09298215.2018.1515233
30. Wallmark Z (2013) Big data and musicology: new methods, new questions. American musicological society national meeting, Pittsburgh, PA. Technical report
31. Weiß C, Mauch M, Dixon S, Müller M (2018) Investigating style evolution of western classical music: a computational approach. Musicae Scientiae
32. Weyde T, Cottrell S, Dykes J, Benetos E, Wolff D, Tidhar D, Kachkaev A, Plumbley M, Dixon S, Barthet M, Gold N, Abdallah S, Alancar-Brayner A, Mahey M, Tovell A (2014) Big data for musicology. In: Proceedings International Workshop on Digital Libraries for Musicology, New York, USA

Experimental Investigations and Future Possibilities in Network-Mediated Folk Music Performance

Chrisoula Alexandraki

Abstract This chapter is intended to acquaint music ethnologists with the paradigm of Networked Music Performance (NMP). NMP facilitates computer networks to allow musicians from distant geographic locations to synchronously collaborate during performance, improvisation or more generally music-making. The chapter comprises two parts. The first part is devoted to providing an overview of research approaches in NMP and elaborates on the technical and perceptual impediments restricting the wide availability of this type of technology. The second part presents an experiment involving three musicians performing folk music over the network. The experiment serves to reveal not only technical and perceptual difficulties in the communication of performers, but more importantly their attitude towards engaging in this novel practice. The chapter concludes by discussing future perspectives on the use of NMP technology in the context of ethnic and folk music.

1 Introduction

Nowadays, network communications are used for almost every form of daily activity. In music and musicology, this abundance of networking applications demands for reassessing conventional practices not only for music preservation and distribution, but also for music-making.

In line with this reasoning, the adoption of digital audio recordings in music preservation has offered significant improvements not only in terms of the quality of the recorded music, but also in terms of documenting and maintaining relevant information in music archives. Furthermore, the recent proliferation of cloud computing technologies has enhanced the visibility of music archives as well as of pertinent research instruments. Yet in more recent days, advances in semantic technologies and computational intelligence allows redefining the role of music archives from isolated

C. Alexandraki (✉)
Department of Music Technology and Acoustics Engineering, School of Applied Sciences, Technological Educational Institute of Crete, Heraklion, Greece
e-mail: chrisoula@staff.teicrete.gr

R. Bader (ed.), *Computational Phonogram Archiving*, Current Research in Systematic Musicology 5, https://doi.org/10.1007/978-3-030-02695-0_10

storage repositories to linked data resources, hence promoting novel research perspectives in massive processing of music data including the unveiling of transcultural characteristics and transdisciplinary research possibilities. A potential next step to this progress is on the actual act of music-making, including the improvisations of folk musicians and the musical interactions taking place among different cultures and ethnic groups.

This chapter focuses on NMP technologies and discusses perspectives emerging from their adoption by folk musicians. NMP can be thought of as a distinct type of teleconferencing application, which allows remotely located musicians to engage in synchronous music performance using network infrastructures and dedicated software tools. Teleconferencing and Voice over IP technologies have a history of more than thirty years now and are being broadly used for an abundance of daily collaboration activities. They are possibly the most prominent type of groupware in computer-supported cooperative work. However, as music performance constitutes a special creative activity with severe restrictions in timing synchronization as well as in motor and cognitive engagement of participants, the progress in teleconferencing has not been effectively propagated to the music domain. As a result, the potential of folk musicians using NMP technology to engage in distributed music performance remains to be elucidated.

The rest of this chapter is structured in two parts. The first part presents an overview of NMP research approaches by elaborating on the challenges faced by this line of research and the workarounds or achievements in meeting these challenges. The last section of this part elucidates that depending on the actual scenario or interaction context among musicians, NMP sessions may have a different degree of technical complexity and that there are in fact scenarios in which synchronous collaborations of musicians may be feasible through the network. The second part of the chapter is devoted on describing the experience of synchronous collaborative performance in the absence of co-presence and how it can be realized in the context of folk and traditional music. An experiment of three musicians performing two pieces of the traditional music of Crete is presented along with an evaluation concerning not only technical measurements, but also qualitative aspects delineated by interviewing performers. Finally, the chapter discusses the potential of folk musicians widely adopting NMP technology and the influence of this technology on the appearance of new music styles emanating from cross-boundary music collaborations.

2 Networked Music Performance Research

Since its inception, NMP was not intended to substitute conventional performance. In fact, professional musicians, unless acquainted with the practice of avant-guard music performance, appear sceptical of the idea of being physically separated from their peers. Unlike speech, music collaboration relies on sharing common ambience in terms of sound reverberation as well as in terms of visual communication concentrating on motor interactions and eye contact between musicians. Hence, physical

proximity of musicians and co-location in physical space are typical prerequisites of collaborative music performance. Yet, musicians are captivated by the use of network-mediated music performance not only as an experimental music practice, but also as an enabling practice when physical co-presence is not possible. NMP is considered to be an enabling technology in cases of travelling obligations of musicians or as an opportunity to reach musicians of a distant geographic region.

The idea of music performers collaborating across distance was remarkably intriguing since the early days of computer music research. Early experimental attempts on exploring the aesthetics of network music interaction date back to the to the 70s [1, 2]. However, in these approaches the focus seems to be placed on machine interaction rather than on the absence of co-presence, as in all of these initiatives the performers were in fact collocated and connected through Local Area Networks (LAN). Telepresence across geographical distance initially appeared in the late 1990s [3] either as control data transmission, noticeably using protocols such as the Remote Music Control Protocol (RMCP) [4] and later the OpenSound Control [5], or as one way transmissions from an orchestra to a remotely located audience [6] or to a remote recording studio [7]. True bidirectional audio interactions across geographical distance became possible around 2001 with the advent of academic network infrastructures, specifically the Internet2 in the US and later the European GEANT. These infrastructures offer highly-reliable broadband connections that are necessary for collaborative music performance.

Since then, a number of research projects have been initiated. Despite the almost two decades of a substantial number of research efforts, the main challenges faced by the implementation of NMP applications have not been defeated. Currently, NMP is only feasible on academic networks and under certain limits in geographical distance. On the other hand, widely available network infrastructures are characterized by a number of technical constraints that impede meeting the perceptual prerequisites of musicians during live performance. Hence, NMP facilities are only available to the academia.

At present, NMP research is highly interdisciplinary as it involves numerous aesthetic, technical and perceptual aspects. An extensive overview of research efforts on NMP is beyond the purposes of this chapter as there are dedicated works available in the relevant literature [8, 9]. The following sections attempt to provide an overview of basic concepts, challenges and approaches that will allow interested researches to gain an understanding on the current issues and future perspectives of NMP research.

2.1 Technical and Perceptual Challenges of NMP Research

Undoubtedly, one of the main reasons why NMP remains an unsolved problem is related to what we know of the cognitive processes and the perceptual qualities involved in synchronous collaborative music performance. Unlike conventional music performance and as portrayed in Fig. 1, NMP fosters collaboration of musicians located in dissimilar environments, with respect to several modalities. To understand

Fig. 1 NMP fosters collaboration of musicians located in dissimilar environments, with respect to several modalities

this type of interaction, one needs to understand the cognitive processes that enable musicians to synchronise during conventional ensemble performance. A number of dedicated studies [10–12] confirm the fact that defining metrics and thresholds for successful collaboration of musicians is a poly-parametric and tremendously complex problem.

The next sections discuss the main technical impediments of NMP systems and their relation to perceptual aspects of collaborative music performance.

3 Communication Latency and Latency Tolerance

The most prevailing problem of NMP technology is the communication latency occurring among remotely located performers. This latency is due hardware and software equipment, network infrastructures, and the physical distance separating collaborating musicians. Specifically, in the simple case of performer A transmitting audio signals to a remote peer (performer B), the lifecycle of an audio segment will undergo the following processes:

(a) *Audio acquisition*: Audio gets captured and digitized in small segments by the hardware equipment of performer A (i.e. microphone, sound card).
(b) *Buffering*: These segments are buffered in small blocks of a predefined size.
(c) *Packetization*: Each resulting block is integrated with additional data to form a network packet. This additional data is known as *packet header* and it contains network specific information such as the address of the destination of a network packet.
(d) *Transmission*: Each network packet is propagated to the network though a routing path, which is determined by the instant network traffic, and therefore not known beforehand.
(e) *De-packetization*: The packet is received at its destination and the audio block is retrieved out of the packet
(f) *Playback*: The audio block is queued for playback by the hardware equipment (i.e. soundcard and speakers or headphones) of performer B.

Clearly, this sequence of processes is bidirectional, i.e. not only from performer A to performer B but also from performer B to performer A. Each of the these processes has a different contribution to the entire communication latency. In common setups, the higher contributions of latency occur during process (b) buffering, resulting in what is called *buffering delay* and process (d) transmission, causing *network latency.*

Buffering delay refers to the time required for acquiring an audio segment from the sound card. The length of this segment corresponds to a certain number of samples and therefore a certain time interval, which depends of the sampling rate. So, for example buffering 256 samples of CD quality audio corresponds to (256/44,100) s = 5.8 ms, which, for the purposes of music performance, corresponds to a significant amount of delay.

Network latency refers to the time elapsed for a network packet to reach its destination. The routing path of a network packet is neither known beforehand nor can be controlled. Depending on the actual transmission path, a packet may require a long time to reach its destination, because it may be held up in long queues, or take a less direct route to avoid congestion.

Besides the value of network latency, a more important obstruction in NMP is related to the fact that the different network packets will reach their destination with different delays. Variation in the delivery time of different packets is known as *network jitter.* Network jitter may be due to queuing network packets on different network devices across the transmission path, or due to packets being driven in different routing paths. Since audio playback requires a steady pace, jitter must be eliminated. Reducing jitter in the network requires stable routes, which are generally not feasible on the Internet on an end-to-end basis.

In teleconferencing and VoIP, the total amount of communication latency is known as Mouth-to-ear latency. Speech-based human interaction is highly tolerant to latency, with a threshold of approximately 150 ms [13]. Unluckily, in music performance this threshold is lower by an order of magnitude, which is the main reason why the progress of teleconferencing has not been effortlessly propagated to the music domain.

Since the first years of NMP research, a number of studies are being performed for the purpose of effectively measuring latency tolerance during ensemble performance. For Schuett [14], this objective was defined as identifying an *Ensemble Performance Threshold (EPT)*, or "the level of delay at which effective real-time musical collaboration shifts from possible to impossible". Schuett observed that musicians would start to slow down performance tempo when the communication delay was raised above 30 ms. This value was further confirmed by a number of subsequent studies [15, 16].

Yet, all of these studies acknowledge that the adaptability of musicians to performing with latency depends on various factors such as their music background, their skills and level of proficiency as well as their age and their familiarity with technology in general [17–19]. Besides the profile of musicians, a number studies show that adapting to latency is highly dependent on certain attributes of the music performed. Such attributes are for example the rhythmic structure [20], the tempo [21] and the timbral qualities of the instruments participating in the music ensemble [22, 23].

4 Audio Quality and Network Throughput

Bandwidth availability or *network throughput* refers to the capacity of the network to accommodate certain data rates. Due to varying load from disparate users sharing the same network resources, the bit rate that can be provided to a certain data stream may be too low for real-time audio communication if all data streams get the same scheduling priority. When the load of the network is greater than its capacity can handle, the network becomes congested. Characteristics of a congested network path are queuing delays, packet loss and sometimes the blockage of new connections.

To date, the majority of NMP architectures use uncompressed audio signals for the communication of remotely located performers. The minimum quality of these signals corresponds to the characteristics of CD quality audio, i.e. signals are sampled at a rate of 44.1 kHz with a sample resolution of 16 bits. In the case of monophonic signals, this audio quality results in a constant data rate of 689 kbps per audio channel. Considering network communication, this is less than the actual data rate required by the network, since, as previously discussed, network packets comprise not only audio data but also header information. Moreover, offering improved music collaboration experience would commonly require higher sound quality as well as multiple channels of audio.

To address this problem a number of research initiatives are focusing on compressing audio data prior to packetization and transmission, hence reducing the required network throughput. Despite the reduction of data rate, this optional step of audio encoding increases the total communication latency for two reasons. Firstly, because it introduces an algorithmic delay caused by signal encoding and secondly because, for efficient reduction of data rates, an audio segment of substantial length needs to be acquired, hence increasing the buffering delay. For instance, the Opus codec [24]

which is currently the de facto standard for real-time audio streaming over IP, recommends a buffering delay of 20 ms for full-band monophonic audio and is associated with a processing delay of 5–65 ms. This is rather prohibitive if one considers the approximate value of the EPT (i.e. 30 ms).

As a result, low latency perceptual codes are increasingly taking into account the requirements of NMP systems. Examples of such works are presented by the Soundjack application, which uses the Fraunhofer Ultra-Low Delay (ULD) Codec [25] and the integration of the WavPack codec by researchers at the Technical University of Braunschweig [26].

5 Audio Dropouts and Data Loss

Finally, *packet loss* occurs when one or more packets of data travelling across a computer network fail to reach their destination. In Wide Area Networks (WAN), packet loss is frequently observed and caused by congested network paths or data corruption by faulty networking hardware across the path. In the case of audio, losing network packets will result in audio dropouts at the receiving end. Audio dropouts correspond to signal discontinuities perceived as glitches, which, in the case of NMP, can seriously hinder the collaboration of music performers. To provide a better insight to the type of distortions caused by packet loss, Fig. 2 depicts a piano signal with severe packet loss occurring during transmission over a ADSL network.

In the domain of network technologies, there are several approaches to recovering from packet loss. For instance, for the widely used Transmission Communication Protocol (TCP), in the event of a lost packet, the receiver asks for retransmission or the sender automatically resends any packets that have not been acknowledged by the receiver. This method of error handling is known as Automatic Repeat reQuest (ARQ). Clearly, ARQ is not an appropriate error correction method for real-time multimedia communications, as in teleconferencing or NMP the packets received after retransmission will be outdated.

Alternative methods for recovering from packet loss include *error concealment* [27] and *Forward Error Correction* (FEC) [28]. Error concealment attempts to

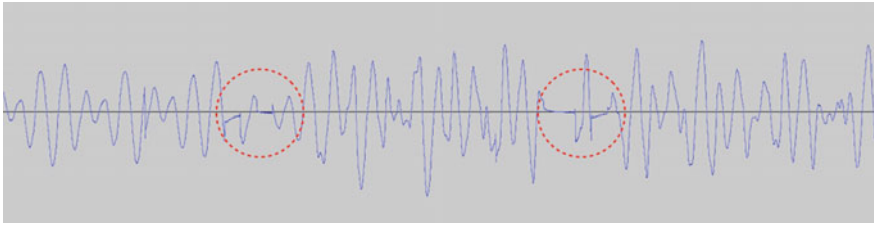


Fig. 2 Heavily distorted piano signal by packet loss during transmission, shown by dashed circles, over ADSL network

recover missing signal portions by using signal processing techniques such as interpolation, pattern repetition or silence substitution. Conversely, Forward Error Correction methods transmit redundant information in addition to actual data packets and attempt to recover losses by reading this redundant information. Information redundancy may be systematic, if it is a verbatim copy of the original data, or non-systematic, if it represents some code that can be facilitated to recover the original data. Clearly, error concealment has the drawback of adding a certain amount of algorithmic latency while FEC methods have the drawback of increasing the requirements in bandwidth availability.

6 Lack of Immersion and Relevant Interaction Affordances

Evidently, communication during music performance does not simply account to musicians listening one another. Performers rely on multiple interaction modalities to synchronise and efficiently communicate. Qualitative studies concerning the requirements of musicians in the absence of co-presence [29], have shown that in addition to sound, visual interactions are necessary for effective synchronization. For this reason, most NMP frameworks employ real-time video, in addition to audio streams, hence further increasing the requirements in network throughput. Video data rates are normally higher than audio ones. However as in music performance, vision is less critical than sound with respect to latency and quality, video communication allows for applying extensive data compression. State of the art NMP frameworks have been experimenting with the use of MJPEG [30], MPEG4 [31] and H.263 [32] video codecs. Evaluation experiments [30] of NMP systems offering video communication in addition to audio, revealed that the main problem in the video communication of performers was the range of viewing angle, rather than the performance of the video codecs. This fact suggests that motor interactions are significant during performance and therefore the positioning of cameras must be carefully considered. Moreover, the use of a data projector as an alternative to the computer monitor can further improve the comfort of musicians.

Besides effective video communication, the use of immersive audio has been considered as an enhancement to performers' feeling of immersion. Specifically, in the work of Chafe [33], a Distributed Internet Reverbarator for Audio Collaboration (DIRAC) was implemented using comb filters to provide the illusion of performers sharing common room reflections of the audio signals reproduced in different locations. The Distributed Immersive Performance (DIP) experiments (Sawchuck et al. [31] used a 10.2 channel immersive audio system so as to represent temporally and spatially distributed audio cues that were created by the interactions of each sound source with the acoustical elements of the environment. Finally, in the MusiNet project the use of spatial audio techniques was considered as an attempt to render the spatial attributes of the audio scene of each performer, thus achieving a more realistic acoustic sensation [34].

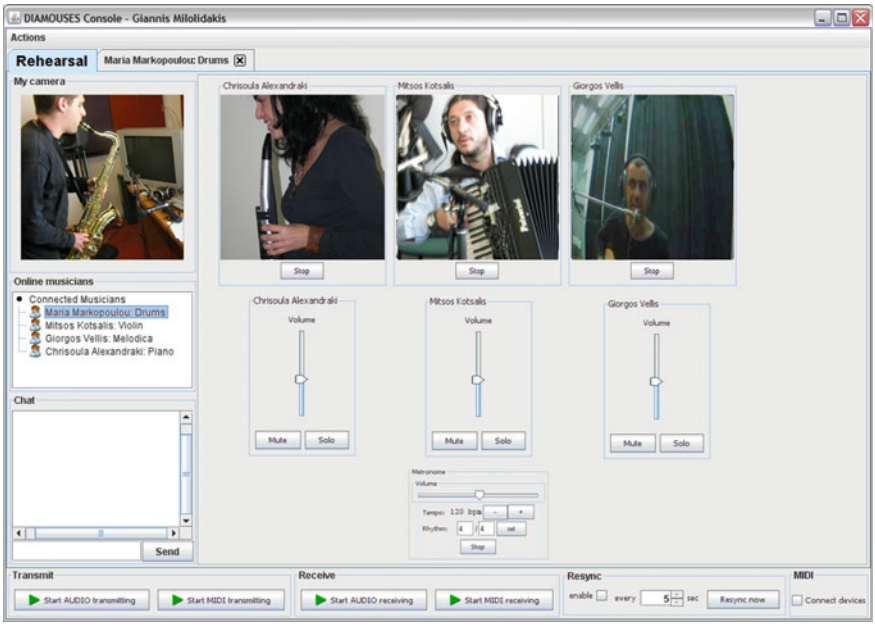


Fig. 3 A GUI developed for the purposes of a networked music rehearsal in the context of the DIAMOUSES project

Yet, a further aspect of research concerns the study of man machine interfaces that can efficiently replicate interaction practices employed by musicians during music performance. Such practices include for example the use of music notation, the use of a metronome or the presence of a conductor. The requirements for employing such concepts or artifacts depend on several defining characteristics of the music performance context. As further elaborated in the section that follows, these characteristics may for example be the kind of music being performed or the purpose of a music performance session. Clearly, different contexts of use raise different requirements in terms of the interaction practices that must be supported. For instance, in remote music-learning scenarios the research focus is on supporting appropriate pedagogical paradigms and on exploring methods for the evaluation of student progress [35]. Alternatively, in network-mediated collaborative music composition [36], a key challenge relates to representing musical events effectively and devising appropriate musical notations [37]. In the context of the DIAMOUSES framework [30], research investigations focused on accommodating diverse user requirements in music performance across different collaboration scenarios such as rehearsals, stage performances and music lessons. Figure 3 depicts an instance of the Graphical User Interface (GUI) developed for supporting distributed music rehearsals. Besides video communication, this GUI provides a chat facility, a metronome as well as the possibility to control the audio levels of individual peers.

Such interfaces integrate the audiovisual communication of musicians with shared collaborative objects permitting synchronous manipulations accessible to all participants, hence maintaining a sense of user focus and promoting a collaborative perspective. Computer Supported Collaborative Work (CSCW) is a focal point of research for several application domains, including e-gaming, e-learning and enterprise groupware, to name a few. In the domain of network-mediated music performance, this perspective has not been adequately investigated.

6.1 Challenging and Feasible NMP Setups

In agreement with the discussion of the previous section, it becomes apparent that the context of music performance severely affects not only the interaction practices employed among performers, but also their requirements in the quality of audiovisual communication. The efficiency of an NMP system to support communication and interaction throughout a networked music session, depends on various parameters that are portrayed in Fig. 4. Four independent parameters determine the feasibility, the requirements and the efficiency of a NMP setup. These are the purpose of the performance session, the characteristics of the music performed, that may be summarized as the music genre, the type of the communication network facilitated and the supported modalities in the communication of performers. Selecting different values for each of these parameters can make up more or less feasible NMP scenarios with varied degrees of complexity.

Fig. 4 Parameters determining the feasibility of a networked music performance session

Out of these parameters, the most prevailing is the purpose of the performance, as it determines the requirements in synchronisation and the required quality of audio-visual communication. Specifically, in the course of a music lesson or a masterclass, musicians will rarely perform at the same time. Usually, the instructor dictates a rhythmic or melodic pattern to be reproduced by the student or the audience of a class. Thus latency is not crucial to the outcome of the session. In contrast, rehearsal or jamming sessions, require a great amount of synchronisation and are highly intolerant to latency. Further possibilities include distributed stage performances and remote recording sessions [7]. In the case of live stage performances requirements will be different depending on whether musicians collaborate from distance, in which case low latency and high fidelity is critical or whether the performance of an collocated ensemble is transmitted to a remote audience. Although, audience feedback forms an essential stimulus for performers, this second scenario is much more feasible as the communication in the direction of the audience to the performers is neither time- nor quality-critical. Hence, this setup is highly feasible and it represents what is widely known as live streaming.

The success of a NMP session is largely dependent on intrinsic attributes of the music performed (e.g. melody rhythm, instrumentation). This was previously discussed as a determining factor to musicians' latency tolerance. Recently, there appears to be debate on whether music genre can reflect any of these attributes [38, 39]. Such considerations have emerged from the recent proliferation of novel music services and applications such as music streaming applications (e.g. Spotify, Last.fm, Pandora) as well as of computational models for automatic genre classification [40]. Consequently, the above schema is susceptible to a great amount of critique.

However, even though musical genre may not readily translate to musical structure or musical style, when considering the act of performance it does translate to the interactions taking place among musicians (e.g. use of notation, leading/accompanying performer intertwining, rhythm interlocking etc.) as well as the occasion for which performance takes place (e.g. festival, celebration etc.). Moreover, music genre may roughly describe the profile of musicians and hence their willingness or motivation to engage in network music performances. With respect to the occasion of performance and the profile of musicians, the two extremes are possibly ethnic/folk music in which there is commonly no scheduling of performance and electronic/electroacoustic performance, which, due to its experimental nature and the intrinsic use of technology, can creatively accommodate network deficiencies, as for example done by Cáceres and Renaud [41] for network delays.

A further parameter of Fig. 4, namely the modalities supported by the NMP session determine the required availability of network throughput. Audio and video streams are more demanding in terms of data rates, while gesture and control data protocols (i.e. those controlling electronic instruments like MIDI or Open SoundControl) are lighter with respect to their data rates.

Finally, the network infrastructure determines the technical efficiency that can be provided during NMP, as well as the geographic distribution of performers. Although the idea of NMP implies that performers are geographically distributed, most NMP research experiments are performed using a Local Area Network in which performers

are distributed in different rooms or buildings sharing the same LAN. This is due to the fact that LAN is more reliable in terms of latency and bandwidth and it practically exhibits zero packet loss. Hence, LAN is more appropriate for studying the attitude of performers towards being physically separated. Moreover, LAN allows for artificially inducing adjustable amounts of latency and packet loss either coded in the NMP software [15] or via using network simulation software such as GNS3 (https://www.gns3.com/) [32], therefore investigating the tolerance of musicians to various network conditions.

6.2 Assessing NMP Experience

The experience of participating in NMP experiments is difficult to describe. My personal involvement in NMP research began in 2005. Since then, I have participated in two funded research projects. The *DIAMOUSES* (www.teicrete.gr/diamouses) project that took place during the years 2006–2008 and the *MusiNet* project (http://musinet.aueb.gr/) that took place between 2012 and 2015. Both of these projects resulted in the development of technological frameworks enabling NMP. In DIAMOUSES, developments were based on the assumption that technology will eventually prove feasible for certain NMP contexts. Thus, the main focus of the DIAMOUSES framework was to support collaboration during different NMP scenarios (e.g. lesson, jamming, live performance), hence fostering the formation of virtual music communities. MusiNet had a different research orientation. In this framework the focus was to develop a user-friendly 'Skype-like' approach for music. Therefore developments were based on state of the art low-latency codecs for audio and video compression, as well as on network protocols and architectures that can provide affordances such as presence awareness (i.e. indication of the status of the user e.g. online, away, busy) and conference calls. Providing further technical details on the development of these frameworks is far beyond the purposes of the present chapter and may be found in dedicated publications [30, 32].

During these years, a number of evaluation experiments were conducted by inviting musicians of different backgrounds to experiment with the frameworks under development. These experiments spanned several combinations of the parameters depicted in Fig. 4 and they had a dual purpose. On one hand to evaluate technical performance of the frameworks under development, the so called *objective evaluation* and on the other hand to evaluate the experience of musicians using these frameworks, namely the *subjective evaluation*.

6.3 Folk Music Experiment

Of the experiments carried out during my involvement in NMP research, the most relevant to this chapter is the Folk music experiment, conducted in the context of

the MusiNet project. The setup involves three musicians performing two pieces of the traditional music of Crete. The following sections describe the pilot setup and the evaluation results, presented as objective measurements and subjective qualities. The findings are inspiring for discussions on the adoption of NMP technology by different cultures, its capacity to foster cross-cultural collaborations and its future role on the development of new music.

7 Pilot Setup

The experiment was conducted in October 2015 in the city of Rethymnon at the premises of *Department of Music Technology and Acoustics Engineering* of the *Technological Educational Institute of Crete*. Table 1 depicts the performers' instruments and their music background. All three instruments are string instruments that are typical for the traditional music of Crete. Musicians performed two pieces, namely *Protos Syrtos* or *Chaniotikos,* which is a traditional dance piece with the rhythm of 2/4 and *Paradosiakes Kondylies*, which is an improvisation on melodic patterns of Crete performed in 4/4.

The performers were initially situated in the same room to agree upon their performance, as shown on Fig. 5. Subsequently, they were asked to move to different buildings of the Department campus, where the MusiNet client equipment had been setup. The framework was configured to allow communication through LAN using a streaming server that was developed for the purposes of the MusiNet project [32, 34]. This server was receiving the audio and video streams from each musician, which were relayed to the remaining two performers using the *Realtime Transport Protocol* (RTP), which is a protocol that is typically used for media telecommunications.

Table 1 The profile of performers that participated in the Folk music experiment

Musician	Instrument	Music background
Alexandos	laoúto (λαούτο)	He says that he comes from a 'musical family' and he sings since he was a little boy. He plays the piano for ten years and he has a certificate in Byzantine music, theory and harmony
Stratis	Mandolin	He holds a degree in Byzantine music, he is a self-taught musician of traditional Cretan music and he works as a professional musician
Minas	lýra (λύρα)	He is a self-taught musician of traditional Cretan music and he works as a professional musician

Fig. 5 Musicians performing at the same site prior to the NMP experiment. The instruments in the order of left to right are Mandolin, Cretan Lyra and Laouto

During network-mediated performance, the hardware equipment of each performer comprised a microphone, a pair of headphones, a camera and a high fidelity sound card connected to a Mac OSX computer which was running the MusiNet client software. As depicted on Fig. 6, the client software provided a GUI which was displaying the video feeds from the from the other two performers as well as a self-view that permitted proper positioning of the local camera. Audio quality was set to 48 kHz, 32 bit, mono, compressed using the Opus codec and captured/transmitted with a buffering delay of 5 ms. The captured video had a 352×288 pixels display resolution, a frame rate of 25 fps and it was encoded using the H.263 codec. According to these quality characteristics and if no media compression would be applied, the data rates would be 1.46 Mbps for each audio stream and 2.42 Mbps for each video stream. As each MusiNet client was transmitting one audio and one video stream to the server and receiving two audio and two video streams, a minimum of 3.88 Mbps upload rate and a 7.76 Mbps download rate was required for the communication of the three performers, excluding the overhead of the headers of network packets.

Fig. 6 Stratis performing the mandolin and communicating using the MusiNet pilot setup

8 Evaluation

To evaluate the technical efficiency of the MusiNet network in supporting NMP sessions, a number of measurements were captured using a dedicated network analyser software, called *Wireshark* (www.wireshark.org). Wireshark permitted capturing the upload and download network traffic at the location of each network node (i.e. each performer as well as at the location of the MusiNet server). Analysis of this traffic permitted the estimation of average values for latency, data rate and packet loss. These averages are shown on Table 2.

In network experiments, measuring data rates and packet loss is pretty straightforward, as the software estimates the number and size of transmitted and received packets. The most challenging task is to measure network latency and this is due to the fact that computers participating in the communication are not synchronized with the required accuracy which needs to be of order of few milliseconds. One workaround to this problem is to use the *Network Time Protocol* (NTP) to synchronise different computers to the same time.

Using NTP to measure latency, the average value of latency shown on Table 2 corresponds to the total time elapsed after capturing 5 ms of audio, encoding, trans-

Table 2 Average values depicting network traffic during the folk music experiment

	Audio	Video
Latency (ms)	9.3	7.3
Outbound bandwidth (kbps)	17.92	81.06
Packet loss (packets)	0	0

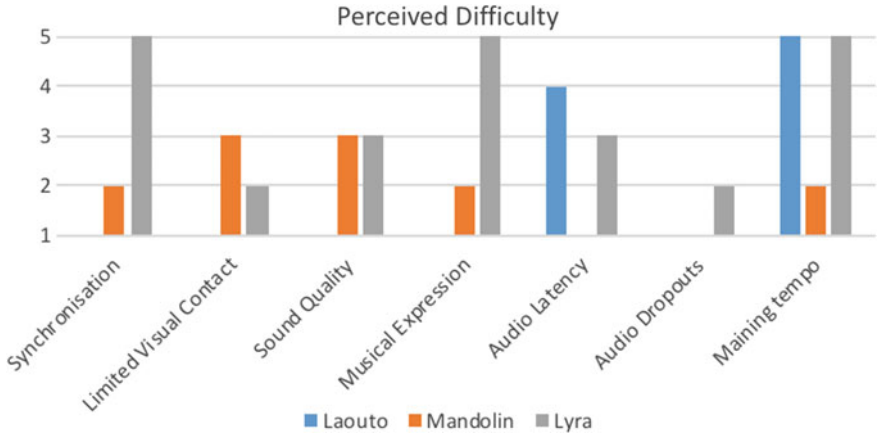


Fig. 7 The perceived difficult of musicians during network-mediated performance. Different aspects of communication were rated in a scale of 1–5, with 1 denoting no difficulty

mitting, decoding and reproducing it to the receiver. Therefore it includes all processing and network delays, apart from the buffering delay, which is 5 ms. The value of 9.3 ms is one third of the approximate EPT value of 30 ms, and should therefore be negligible by performers. As for bandwidth, Table 2 depicts the audio and video rates at the direction of transmission (i.e. the outbound network traffic). These rates correspond compression ratios of 30:1 and 83:1 for audio and video respectively compared to the raw, uncompressed rates that were mentioned earlier. Finally, no packet loss was observed, which was anticipated due to the fact that the experiment was based on a properly configured LAN.

To assess the subjective qualities of the experiment, we interviewed performers and asked them to fill in a questionnaire. Among other questions, the questionnaire asked them to rate the perceived difficulty with respect to different aspects of interaction during performance. This rating is depicted in Fig. 7. It appears that their main difficulty was on maintaining the tempo during performance and, according to the interview, this was particularly perceived when performers attempted to intentionally perform with a very slow or a very fast tempo. They felt more comfortable when performing with a moderate tempo. They reported having synchronisation problems and that they were able to perceive ‘some latency’. Limited visual communication was reported as having a moderate influence on their performance. This can be attributed to the fact that, as video and audio were not precisely synchronised led to participants ignoring the video feeds during performance. Finally, musical expression was problematic, especially for the performer that is less acquainted with audio technology, i.e. Minas performing the lyra, who reported that technology limits freedom in folk music performance. He also mentioned that the amplitude levels of his co-performers were problematic, which made him feel uncomfortable.

To conclude this experiment, it appears that even in ideal network conditions (e.g. those available on LAN) the communication of performers is not adequately

efficient to effortlessly accommodate the 'fine-grained' audio-visual interactions of musicians. However, despite the fact that performers were not totally comfortable with the provided setup, they reported that they are whiling to adapt, especially when they are unable to meet. In fact, one of these performers, namely Alexandros performing Laouto, expressed a high interest in using this system to teach Cretan music to Greek expatriates. Although, as previously discussed, teaching is relatively more feasible to support, we informed Alexandros that using the MusiNet setup on a commonly available network, would result in high percentages of packet loss perceived as excessive signal distortion. Consequently, his desirable scenario would only be achievable, either by introducing error concealment techniques, hence further increasing the communication latency or over a highly reliable network infrastructure, which can only be provided in academic setups. In conclusion, we suggested that using the MusiNet system to teach Greek expatriates on a common network infrastructure would require a configuration that would not offer any reward to using existing teleconferencing solutions (e.g. Skype).

9 Discussion

Besides technical observations and feasibility considerations, the presented experiment provided insight on the attitude of folk musicians towards the novel practice of performing music while being physically separated. Although professional folk musicians may be familiar with audio technology, for example in course of stage performances or recording sessions, when it comes to remote collaborative improvisation they are sceptic of the efficiency of technology to support music expression. Moreover, as ethnic music typically appears to be performed in the occasion of another activity such as working, eating, drinking or in social events such as weddings, funerals, celebrations, it is difficult to envisage NMP being widely adopted by performers of this style of music. Nevertheless, traditional music should not be considered stagnant, as it does and will continue to embrace technological developments in a myriad of ways.

It is widely acknowledge that today, technological advancements are rapidly embraced by the public at large and used in our daily activities. This is also true for music and music information not only for consumers but also for music artists. Musician-blogs, social networking of musicians and indie-artist friendly streaming services have provided a new paradigm to artist promotion. With respect to ethnic music, this fact has led to the emergence of new genres as well as a constructive criticism for the validity of the existing ones [38, 39]. The genre of *world music*, although heavily criticized for its use by the music industry [42], reflects a global trend for the creation of controversial musical sounds by amalgamating music constructs originating from diverse geographical regions. In the last century, world music, fusion and hybrid genres appear to be one way to foster music controversy, possibly the other one being the development of electroacoustic music. The ever increasing appeal of music artists to these two genres, appears to evolves in parallel with technological

developments. Besides electroacoustic music, which is inherently based on technology, the growth of world music styles can largely be attributed to the fact that musicians from diverse cultures can easily access high-fidelity samples of ethnic music, sound bites and loops as well as the fact that musicians can easily reach one another by means of social networking platforms.

A reasonable extension to this trend can be realized by the advancement of NMP technology. NMP, compared to artist promotion and social networking, provides a dimension of real-time synchronous interactions, hence fostering the immediate and imminent development of new music styles. As already discussed in the introductory section, NMP falls in the category of new technological developments that have not yet been realized as a widely common practice. Indeed, this due to the various technical constraints that were discussed in the previous sections. However, as network infrastructures become increasingly efficient and as more and more research is devoted on improving NMP technology, it is reasonable to expect that at some point in the near future, NMP technologies will be adopted by music communities at large, hence inspiring cross-cultural music collaborations and promoting the fruition of new music.

With respect to improving NMP technology, a distinct line of recent research efforts concentrates on devising computationally intelligent approaches that can allow for overcoming the main technical impediment of NMP, that of latency. Audio signals traversing large geographical distances will always exhibit substantial communication latency, even with the most sophisticated future networks reaching the speed of light. Consequently, a natural workaround to this problem seems to be in predicting music performance ahead of time and rendering the predicted performance to the remote site at a steady pace and before the actual signal reaches its destination [43]. This idea is closely related to a cognitive phenomenon of music perception known as *music anticipation*. Anticipation is a fundamental characteristic of ensemble performance and it refers to the fact that when the members of an ensemble know each other's performing style very well, they know exactly when their peer will play a note in advance, so before the note is actually played. This type of intelligence emanates from different knowledge processes including the cognitive understanding of a performance plan (e.g. the music score or any alternative construct of pre-existing arrangement), the build-up of the music piece up to that time and finally the experienced gained through past rehearsals of the music ensemble. Therefore, one can develop computational models that can be trained according to these knowledge processes. This approach is employed in an alternative research domain, that of *computer accompaniment systems*. There, the objective is to develop intelligent computer agents that are able to replace any of the members of an ensemble performance [44, 45].

Possibly the first work exploiting computer accompaniment systems to enable synchronous networked performance was the development of a system called *TablaNet* [46]. TablaNet was a real-time online musical collaboration system for the *tabla*, a pair of North Indian hand drums. Hence, this system was in fact dealing with Indian ethnic music performed over the network. The two drums produce twelve pitched and unpitched sounds called *bols*. The system recognizes the performed bols and

the recognized bols are sent as symbols over the network. A computer at the receiving end identifies the musical structure from the incoming sequence of symbols by mapping them dynamically to known musical constructs. To cope with transmission delays, the receiver predicts the next events by analyzing previous patterns before receiving the original events. The predicted events are synthesized by triggering the playback of pre-recorded samples.

More recent initiatives on predicting sound events during performance and rendering prerecorded signal segments at remote network locations, used score following techniques for monophonic acoustic instruments [47]. Most forms of ethnic music performance employ a great amount of anticipation. For example, monophonic vocal music or percussion music of different ethnic groups may be highly suitable for making short-term predictions and can therefore provide the ground for experimental research on anticipatory models. Ultimately, ethnic music may be influenced by NMP technology as well as provide the basis for the advancement of new achievements.

10 Summary and Concluding Remarks

This chapter attempted to provide insight into current NMP technology and its potential use in ethnic and folk music performance.

The first part discussed the main technical constraints that impede meeting the perceptual prerequisites of musicians during live performance. These constraints are the focal points of NMP research and concern: (a) the minimisation of communication latency which leads to synchronisation problems among performers, (b) the high requirements in network throughput that are necessary to provide high-quality audio and video communications, (c) the elimination of network packet loss which results in signal distortions severely hindering the communication of musicians and (d) the lack of immersion in common ambience with respect to various interaction modalities.

Despite constraints, the chapter elaborates on the fact that for different NMP contexts, these restrictions may be more or less important and that there are certain interaction contexts in which NMP collaboration may be highly feasible. For instance, teleconferencing in music learning is less sensitive to communication latencies, as teacher and student rarely perform at the same time and therefore need not synchronise their performance. Equivalently, remote music recording requires high quality in one direction only, while electroacoustic music may be performed with the use of low data rate control signals. Nevertheless, realistic bidirectional music interactions of acoustic instruments, as for example rehearsals, distributed stage performances or jamming sessions in jazz, rock, classic or folk music are only feasible within short geographical distance and over highly-reliable network infrastructures. Currently, the infrastructures that appear to be more appropriate for NMP are those available to the academia. Hence, true, realistic, bidirectional NMP is restricted to academic environments.

To further inform on the actual practice of network-based music performance and how it can be realised in folk music, the second part of the chapter presented an experiment of three musicians performing pieces from the traditional music of Crete through a Local Area Network. As LANs are private networks that span short geographical distances, they are highly efficient in terms of latency, throughput and packet loss. Although they do not represent realistic remote interaction scenarios (i.e. musicians are located in near proximity), they are highly appropriate for experimental research. This is due to the fact that they allow for studying the attitude of performers towards being physically separated and that, if needed, they permit introducing adjustable amounts of latency, throughput and packet loss therefore monitoring the relevant perceptual thresholds.

The presented experiment revealed that folk musicians were not fully satisfied with the quality of their interactions, however they expressed their interest in compromising with this new technology so as to reach a wider audience, namely teach traditional music to Greek expatriates. Overall, the experience gained through the experiment inspires discussions on cross-cultural music collaborations, enabled by the use of NMP technology. As new technological advancements are rapidly diffused in our everyday lives, feasible NMP scenarios are expected to have a significant impact not only on the dissemination of indigenous music, but also on the emergence of new music resulting from broader cross-cultural collaborations.

Acknowledgements I would like to especially thank Alexandros Aggelakis, Eustratios Gounakis and Minas Sfakianakis for volunteering to participate in the folk music experiment. Part of this research has been co-financed by the European Union (European Social Fund—ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF)—Research Funding Program: THALIS–MusiNet.

References

1. Barbosa Á (2003) Displaced soundscapes: a survey of network systems for music and sonic art creation. Leonardo Music J 13:53–59. https://doi.org/10.1162/096112104322750791
2. Follmer G (2005) Electronic, aesthetic and social factors in Net music. Organis Sound 10:185–192. https://doi.org/10.1017/S1355771805000920
3. Kapur A, Wang G, Cook PR, Davidson P (2005) Interactive network performance: a dream worth dreaming? Organis Sound 10:209–219. https://doi.org/10.1017/S1355771805000956
4. Goto M, Neyama R, Muraoka Y (1997) RMCP: remote music control protocol—design and interactive network performance applications. In: Proceedings of the 1997 international computer music conference, Thessaloniki, Hellas, ICMA, pp 446–449
5. Wright M, Freed A (1997) Open sound control: a new protocol for communicating with sound synthesizers. Proc ICMC 1997:101–104
6. Xu A, Woszczyk W, Settel Z, Pennycook B, Rowe R, Galanter P, Bary J, Martin G, Corey J, Cooperstock J (2000) Real time streaming of multi-channel audio data through the internet. J Audio Eng Soc 48(7/8):627–641
7. Cooperstock JR, Spackman SP (2001) The recording studio that spanned a continent. In: Proceedings of 1st international conference on WEB delivering of music, WEDELMUSIC 2001. Institute of Electrical and Electronics Engineers Inc., pp 161–167. https://doi.org/10.1109/wdm.2001.990172

8. Gabrielli L, Squartini S (2015) Wireless networked music performance, wireless networked music performance. https://doi.org/10.1007/978-981-10-0335-6
9. Rottondi C, Chafe C, Allocchio C, Sarti A (2016) An overview on networked music performance technologies. IEEE Access 4:8823–8843. https://doi.org/10.1109/ACCESS.2016.2628440
10. Goebl W, Palmer C (2009) Synchronization of timing and motion among performing musicians. Music Percept 26(5):427–438
11. Keller P (2007) Musical ensemble synchronisation. In: Proceedings of the international conference on music communication science, pp 80–83
12. Rasch RA (1988) Timing and synchronization in ensemble performance. In: Sloboda JA (ed) Generative processes in music: the psychology of performance, improvisation and composition. Clarendon Press, Oxford, pp 70–90
13. Wu X, Dhara KK, Krishnaswamy V (2007) Enhancing application-layer multicast for P2P conferencing. In: Proceedings of the 4th IEEE consumer communications and networking conference, pp 986–990
14. Schuett N (2002) The effects of latency on ensemble performance. Honors Thesis
15. Chafe C, Gurevich M, Leslie G, Tyan S (2004) Effect of time delay on ensemble accuracy. In: Proceedings of the international symposium on musical acoustics, pp 3–6
16. Driessen PF, Darcie TE, Pillay B (2011) The effects of network delay on tempo in musical performance. Comput Music J 35:76–89. https://doi.org/10.1162/COMJ_a_00041
17. Farner S, Solvang A, Asbjørn S, Svensson UP (2009) Ensemble hand-clapping experiments under the influence of delay and various acoustic environments. AES J Audio Eng Soc 57:1028–1041
18. Bartlette C, Bocko M (2006) Effect of network latency on interactive musical performance. Music Percep 24:49–62. https://doi.org/10.1525/mp.2006.24.1.49
19. Chew E, Sawchuk A, Tanoue C, Zimmermann R (2005) Segmental tempo analysis of performances in user-centered experiments in the distributed immersive performance project. In: SMC conference, p 28
20. Rottondi C, Buccoli M, Zanoni M, Garao D, Verticale G, Sarti A (2015) Feature-based analysis of the effects of packet delay on networked musical interactions. AES J Audio Eng Soc 63:864–875. https://doi.org/10.17743/jaes.2015.0074
21. Carôt A, Werner C, Fischinger T (2009) Towards a comprehensive cognitive analysis of delay-influenced rhythmical interaction. In: Proceedings of international computer music conference. http://hdl.handle.net/2027/spo.bbp2372.2009.107
22. Barbosa Á, Cordeiro J (2011) The influence of perceptual attack times in networked music performance. In: Proceedings of 44th international conference: audio networking, 2011, p 10. http://www.aes.org/e-lib/browse.cfm?elib=16133
23. Mäki-Patola T (2005) Musical effects of latency. Swomen Musiikintutkijoiden 9:82–85
24. Valin, J.-M., Maxwell, G., Terriberry, T.B., Vos, K., 2013. High-Quality, Low-Delay Music Coding in the Opus Codec. 135th AES Convention 73–82
25. Kraemer U, Hirschfeld J, Schuller G, Wabnik S, Carôt A, Werner C (2007) Network music performance with ultra-low-delay audio coding under unreliable network conditions. In: Proceedings of the 123rd audio engineering society convention. New York, Curran Associates, pp 338–348
26. Kurtisi Z, Wolf L (2008) Using WavPack for real-time audio coding in interactive applications, in: 2008 IEEE International Conference on Multimedia and Expo, ICME 2008 - Proceedings. pp. 1381–1384. https://doi.org/10.1109/icme.2008.4607701
27. Tatlas N-A, Floros A, Zarouchas T, Mourjopoulos J (2007) Perceptually-optimized error concealment for audio over WLANs. Mediterranean J Electron Commun 3:77–86
28. Xiao J, Tammam T, Chunyu L, Zhao Y (2011) Real-time forward error correction for video transmission. In: 2011 visual communications and image processing (VCIP). IEEE
29. Alexandraki C, Kalantzis I (2007) Requirements and application scenarios in the context of network based music collaboration. In: Proceedings of the AXMEDIS 2007 conference. Florence: Firenze University Press, pp 39–46

30. Alexandraki C, Akoumianakis D (2010) Exploring new perspectives in network music performance: the DIAMOUSES framework. Comput Music J 34:66–83. https://doi.org/10.1162/comj.2010.34.2.66
31. Sawchuk AA, Chew E, Zimmermann R, Papadopoulos C, Kyriakakis C (2003) From remote media immersion to distributed immersive performance. In: Proceedings of the ACM SIGMM 2003 on workshop on experiential telepresence. New York, ACM Press, pp 110–120
32. Akoumianakis D, Alexandraki C, Alexiou V, Anagnostopoulou C, Eleftheriadis A, Lalioti V, Mastorakis Y, Modas A, Mouchtaris A, Pavlidi D, Polyzos GC, Tsakalides P, Xylomenos G, Zervas P (2016) The MusiNet project: addressing the challenges in networked music performance systems, In: IISA 2015—6th international conference on information, intelligence, systems and applications. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/iisa.2015.7388002
33. Chafe C (2003) Distributed internet reverberation for audio collaboration. In: Proceedings of the 24th AES international conference, pp 13–19
34. Akoumianakis D, Alexandraki C, Alexiou V, Anagnostopoulou C, Eleftheriadis A, Lalioti V, Mouchtaris A, Pavlidi D, Polyzos GC, Tsakalides P, Xylomenos G, Zervas P (2014) The MusiNet project: towards unraveling the full potential of networked music performance systems. In: IISA 2014—5th international conference on information, intelligence, systems and applications. IEEE Computer Society. https://doi.org/10.1109/iisa.2014.6878779
35. Ng K, Nesi P (2008) I-Maestro framework and interactive multimedia tools for technology-enhanced learning and teaching for music. In: Proceeding—fourth international conference on automated solutions for cross media content and multi-channel distribution, Axmedis 2008. Florence: Firenze University Press, pp 266–269
36. Hajdu G (2005) Quintet.net: an environment for composing and performing music on the internet. Leonardo Music J 38(1):23–30
37. Hajdu G (2006) Automatic composition and notation in network music environments. In: Proceedings of the 2006 sound and music computing conference. Marseille: Centre National de Creation Musicale, pp 109–114
38. Greenberg DM (2016) Musical genres are out of date—but this new system explains why you might like both jazz and hip hop. Economies. http://www.econotimes.com/Musical-genres-are-out-of-date-%E2%80%93-but-this-new-system-explains-why-you-might-like-both-jazz-and-hip-hop-244941
39. Wong J (2011) Visualising music: the problems with genre classification. Masters of Media
40. Ezzaidi H, Bahoura M, Rouat J (2010) Taxonomy of musical genres. In: Proceedings of 5th international conference on signal image technology and internet based systems, SITIS 2009, pp. 228–231. https://doi.org/10.1109/sitis.2009.45
41. Cáceres JP, Renaud A (2008) Playing the network: the use of time delays as musical devices. Proceedings of International Computer Music Conference 244–250
42. Byrne D (1999) Crossing music's borders: 'I hate world music'. The New York Times. https://archive.nytimes.com/query.nytimes.com/gst/fullpage-9901EED8163EF930A35753C1A96F958260.html
43. Alexandraki C (2014) Real-time machine listening and segmental re-synthesis for networked music performance. PhD dissertation, University of Hamburg. http://ediss.sub.uni-hamburg.de/volltexte/2014/7100/
44. Dannenberg R (1984) An online algorithm for real-time accompaniment. In: Proceedings of the 1984 international computer music conference. Computer Music Association, pp 193–198
45. Vercoe BL (1984) The synthetic performer in the context of live performance. In: Proceedings of the 1984 international computer music conference, Paris, pp 199–200
46. Sarkar M, Vercoe B (200) Recognition and prediction in a network music performance system for Indian percussion. In: Proceedings of the 7th international conference on New interfaces for musical expression NIME 07, pp 317–320
47. Alexandraki C, Bader R (2016) Anticipatory networked communications for live musical interactions of acoustic instruments. J New Music Res 45:68–85. https://doi.org/10.1080/09298215.2015.1131990

Requirements and Use Cases for Digital Sound Archives in Ethnomusicology

Jonas Franke

> *No one can predict the ways their collections will be used. Some will become one of the building blocks of cultural and political movements; some will bring alive the voice of a legendary ancestor for an individual; some will stimulate budding musicians, some will soothe the pain of exile, and some will be used for restudies of primary data that may revolutionize approaches to world music.*
>
> Anthony Seeger [25, p. 264]

Abstract In this chapter, results from requirements elicitation activities for an ethnomusicological sound archiving software are summarized. Results derived from user and expert interviews as well as from literature of the sphere of ethnomusicology, computational musicology, archival studies and informatics. Interviews were conducted with stakeholders that are either involved in creation and maintenance of archiving software or who consider utilizing an archive software for private or professional use. Particular emphasis was placed on requirements and use cases that are supported by recent digital technologies. Online publishing, distributed systems and computational analysis of audio content and metadata can enable ethnomusicological sound archives to bring new value to their corpora. Publishing and sharing contents online can extend academic and private uses and computational analysis of archive contents can help to structure, access and maintain archives in novel ways. As a final result of requirements elicitation activities, technology, architecture and features of an archiving software built based off the findings are briefly presented and future tasks and challenges for this specific implementation are outlined.

J. Franke (✉)
Institute for Systematic Musicology, University of Hamburg, Neue Rabenstraße 13, 20354 Hamburg, Germany
e-mail: jonas.franke@uni-hamburg.de

R. Bader (ed.), *Computational Phonogram Archiving*, Current Research in Systematic Musicology 5, https://doi.org/10.1007/978-3-030-02695-0_11

1 Introduction

Since Thomas Edison (1847–1931) invented the phonograph in 1877, scholars from diverse disciplines such as ethnology, musicology and anthropology recorded countless musical performances during their field studies. These recordings reside today in public or private archives around the world and their great scientific and cultural value is acknowledged for example by the *Berlin Phonogramm-Archiv* being accepted for the *UNESCO World Document Heritage List* in 1999 [13].

Today, ethnomusicological sound archives look back at vital history of more than a century but often do not offer an easy way to access recordings [26]. Thus, the question can be posed, how archives can transition towards an interconnected and data-driven environment. Which innovative applications for their valuable assets can be discovered, how can technology support the process of renewal and what requirements need to be satisfied?

As motivational background for this work the *Ethnographic Sound Recordings Archive* (E.S.R.A.) at the *Institute for Systematic Musicology* must be mentioned. E.S.R.A. includes collections of sound recordings that can be traced back to the early 20th century, when the *University of Hamburg* and academic discipline of comparative musicology in Hamburg were yet to be established. Many of those valuable recordings reside on obsolete media carriers that are endangered of falling apart. Thus an urgent need was discovered, transferring recordings into the digital domain. Further on, musics being recorded in past and recent fieldwork activities by associated students and scholars are being digitized, catalogued and will be made publicly accessible.

To develop an appropriate digital sound archive infrastructure, the *Computational Music and Sound Archiving* (CoMSAr) project was initiated. Scholars at CoMSAr are concerned with digitizing and cataloguing sound recordings of E.S.R.A. as well as working on the actual software development. Additionally, computational methods are implemented to analyze and structure collections. In 2013, a basic database application was developed that was revisited in 2017.[1]

The objective of this chapter is to provide a comprehensive collection of requirements artefacts that were discovered during system engineering processes for CoMSAr. Synergies that could come into play between ethnomusicology, digital sound archiving technology and Music Information Retrieval (MIR) should be emphasized and the question should be discussed, what requirements have to be met by a current ethnomusicological sound archive. As a result of requirements elicitation activities, a short introduction to architecture, technology and features of CoMSAr shall be presented.

[1] http://esra.fbkultur.uni-hamburg.de.

2 Requirements Elicitation

Starting point for requirements elicitation activities was the identification of potential requirements sources. As most important requirements source, future users of the archive were recognized.

Furthermore, literature was identified as an important source for requirements. A vital discourse on sound archiving in ethnomusicology as well as publications on sound archives being utilized in computational analysis could be found. Those resources helped to define a clear image of today's purposes, obligations and opportunities of computational sound archives that led to conclusions about requirements. Additionally, detailed information on standards and best practices could be taken into account, published by organizations like IASA.[2]

A third attempt was to review the legacy implementation of the archive, that has been used by a small team since 2013. Taking one step back reviewing the implementation regarding comments from current users as well as knowledge gained from literature also revealed requirements and improvements.

Additionally, other existing sound archive implementations were examined.

To structure requirements artefacts, the following stakeholder groups were defined, clustering potential users and people that are involved in the software's lifecycle:

1. *Archivists* are working on digitizing records and populating the database with metadata
2. *Consumers* want to listen to sounds and discover music
3. *Developers* are in charge of implementing and maintaining the system
4. *Fieldworkers* want to feed their recordings into the archive
5. *MIR Experts* are working on the implementation of algorithms to extract features and to apply data-driven processing to archive assets
6. *Musicians* want to discover music for the purpose of inspiration or want to publish their music to the archive
7. *Scholars* want to access the archive to incorporate music and metadata into their academic work.

Based on those stakeholder groups, elicitation activities such as user and expert interviews were executed.

Interviews were conducted as *unstructured interviews*, including an opening followed by free conversation employing *open questions* as described by Pohl [21]. At the end of each interview the interviewee was encouraged to write one or more user story cards, according to the *Connextra* format.

Once the actual software implementation started, team-members were invited to provide feedback constantly, allowing the team to enter a continuous requirements engineering process as described by Pohl [21]. Therefore feedback on requirements artefacts as well as the current state of development was elicited within informal follow-ups and small presentations of system scopes.

[2] International Association of Sound and Audiovisual Archives.

To manifest requirements of stakeholders, requirements artefacts derived from interviews and literature were documented as collections of goals. In a second step, scenarios were created, describing a series of actions for satisfying goals defined earlier. Functional requirements artefacts were documented as narrative scenarios grouped by stakeholder groups. The documentation of requirements artefacts followed a template inspired by Pohl [21].

3 Results of Requirements Elicitation

As follows, selected requirements artefacts discovered from literature and interviews shall be summarized. To avoid going beyond the scope of this chapter, requirements artefacts such as user stories and results from literature research shall be summarized and discussed in natural language.

3.1 Worldwide Open Access

Worldwide open access to audio content was considered important by stakeholder groups *scholars*, *consumers*, *MIR experts*, *fieldworkers* and *musicians*. User stories identified in interviews that would be supported by worldwide open access included:

- As a consumer, I want to access recordings for entertainment
- As a fieldworker, I want to play back archive's recordings in the field to show musicians what will happen with recordings of their music
- As a musician, I want to explore music for inspiration
- As a scholar, I want to publish my recordings to reference music in academic work.

The importance of overcoming limited access to archives could also be confirmed by an ongoing discourse in various publications on ethnomusicology and sound archiving. It was found, that all sound archives dealing with intangible cultural heritage could potentially play an important role in today's cultural transmission processes. Many archives own valuable recordings for more than a century without enabling easy public access. Cultural artefacts that might have been lost in between generations of mouth-to-mouth transmission can be discovered by the heirs of the original performers if archives allow public access [26]. Seeger points our attention to the fact, that the audience of sound archives must not be limited to academics and that the answer to the question "who is our primary audience" [26, p. 41] is constantly changing.

As an example for the use of archive recordings for musical inspiration, Feld [6] explaines how the young Bosavi people of Papua New Guinea got inspired by their ancestors' music and gave birth to a 'new guitar band music' movement. Members of the movement wanted to learn more about forgotten songs of their parents and grandparents and recordings could be provided by an archive [6].

Also in Western countries, recordings of folk music can play an important role in influencing cultural identities. As an example, during the 1970s a political grassroots movement arose in Norway, protesting against the dominance of the cities. People on the countryside started taking pride again speaking their own local dialects and an urban folk music revival took place. People were encouraged to search for traces of local music and dance. Indeed, they found local traditions that were almost forgotten. Ethnomusicologists documented music and dance for future usage in sound archives [29].

Furthermore, Seeger [26] mentions that communities increasingly reach out to collectors and archives to recover 'forgotten' traditions. Knowing this, the responsibility of collectors could even be extended: not only to open up, but also to reach out to the communities and let them know that recordings of interest exist. Gray [10] confirms this point of view and extends it by the observation, that community members more often strive to create their own documentation of cultural traditions. Since today's recording and playback devices are not only affordable by the wealthy anymore, a lot of oral traditions are being transmitted utilizing sound recordings already. Tradition bearers record their heritage and the younger generation rehearses performances from sound recordings. This circumstance could be supported by the establishment of local archive initiatives [26].

Additionally, Seeger [26] points to the value of archival techniques and practices, that amateurs are often not familiar with. Within sound archive initiatives, fieldworkers and ethnomusicologists could offer technology and training, since they were able to gain experiences for more than a century. Besides transmitting cultural artefacts, sound archives could become valuable resources for gaining knowledge about archival techniques and best practices as well.

Knowing this, the list of user stories that require public access or at least low access barriers shall be extended as follows:

- As a fieldworker, I want to enable online access to recordings to support cultural transmission processes
- As a musician, I want to contribute recordings to public archives to preserve the music
- As a musician, I want to learn about sound recording and archiving techniques to improve my contributions.

Since today connections to the World Wide Web (WWW) can be established from almost everywhere in the world, a web application seems to be the way with lowest barriers to offer worldwide open access. Technically, every recent personal computer or smartphone is capable of browsing the WWW and playing back sounds from a website [30] and a various range of technologies is ready to be used in web development. Furthermore, the Uniform Resource Locator (URL) concept, that is one of the building blocks of web technology, provides a stable way for referencing in scientific publications.

A challenge can be identified in developing an appropriate *Graphical User Interface* (GUI), that works intuitively under all circumstances. Devices that can display websites usually vary in screen size and ratio. Therefore websites need to adapt to different devices and screens.

3.2 Interoperability

In this context interoperability should be understood as the ability of organizations to share and exchange audio content as well as related metadata. The following user stories were brought up by stakeholders:

- As a MIR expert, I want to analyze audio and metadata throughout various archives to optimize my algorithms
- As a scholar, I want to search for recordings throughout various archives to improve results of my research.

In literature, interoperability is perceived as crucial to productivity and success of individual archives:

> Interoperability must be a key component of any metadata strategy: elaborate systems devised independently for one archival repository by a dedicated team will be a recipe for low productivity, high costs and minimal impact [3, p. 21].

What participants for data exchange can one think of? First, interoperability between sound archives could be enabled. If contents from distributed archives could be accessed and aggregated, results of quantitative research in ethnomusicology could be improved. Since a "large-scale comparative study of world music cultures has not been addressed yet" [20, p. 185], making shared corpora of ethnomusicological sound archive's audio recordings available can be considered as of high interest for scholars of ethnomusicology.

Interoperability could also help solve one of the biggest problems in MIR. As pointed out by Porter et al. [22] and Weyde et al. [33], the unavailability of *Big Data* limits the significance of results and prevents machine learning algorithms from being relevant beyond the small training sets of data currently available to scholars. Weyde, Porter et al. refer to the unavailability of music for MIR in general, not specifically ethnomusicological recordings. Since a huge amount of popular music recordings must be at least available in the digital domain, an even bigger need must be expected for ethnomusicological recordings. Proutskova [24] confirms, and points to the fact, that research on non-Western music is underrepresented in the field of MIR, because music data is not available, incomplete or not standardized. Further on she emphasizes the role that ethnomusicological sound archives play and the potential benefit for MIR in general if more recordings could be considered for research tasks.

Besides interoperability between archives, individual archives could contribute to digital cultural heritage platforms such as *Europeana*,[3] *Dariah*,[4] *Dismarc*[5] or the more generic archive.org.[6] It can be expected that providing assets to these public platforms would widen an archive's audience.

Sharing archive contents with other organizations can support the dissemination processes mentioned in Sect. 3.1 and therefore has the potential to increase an archive's relevance and productivity. The list of user stories can be extended as follows:

- As an archivist, I want to contribute to digital heritage platforms to broaden an archive's audience
- As an archivist, I want to share archive contents with partner organizations to increase an archive's productivity.

A way to enable interoperability is offering an Application Programming Interface (API). By utilizing an API, participants can request metadata and references to audio files based on search queries in an automated way. For implementing an API, the OAIPMH (Open Archives Initiative Protocol for Metadata Harvesting)[7] seems applicable, since it is used by an impressive list of libraries, archives and academic institutions around the globe.[8] OAIPMH is a protocol that defines how sets of metadata are requested and retrieved. Participants of OAIPMH transactions are classified as *repositories*, *harvesters* or *aggregators*. A *repository* could be a sound archive providing an API that offers access to its metadata. A *harvester* could be another institution that uses this API to retrieve metadata (e.g. lists of recently added sound recordings or detailed metadata on single recordings). *Aggregators* are entities that harvest metadata from multiple repositories and offer an interface to access aggregated contents. For data exchange a Representational state transfer (RESTful) API that utilizes the Hypertext Transfer Protocol (HTTP) to request and serve archive resources needs to be implemented. Metadata is encoded in Extensible Markup Language (XML). However, OAIPMH doesn't define the detailed contents or structure of a metadata schema, a *lingua franca* of metadata is to be determined by participants of OAIPMH transactions [3].

3.3 Metadata

Ethnographic sound archives usually deal with metadata related to audio recordings. From an ethnomusicological perspective, a comprehensive documentation of recordings can be considered very important, since "music is not a pure sound

[3] http://www.europeana.eu.

[4] http://www.dariah.eu.

[5] http://www.dismarc.org.

[6] http://www.archive.org.

[7] https://www.openarchives.org.

[8] http://www.openarchives.org/Register/BrowseSites (Visited on 04/20/2018).

experience but has to be judged—although not exclusively—within its ceremonial or ritual context" [27, p. 34]. Such metadata usually contains ethnographic data on original recordings (*when was it recorded*, *what does it contain*, *who recorded it*, *where was it recorded*, *who owns the copyright* etc.) and information on the current medium (*format*, *storage location* etc.). From a technical perspective, metadata is primarily used to search and relate archive assets to each other [19]. For an archive's internal operation, a thorough metadata concept was considered crucial by stakeholders. User stories related to metadata included:

- As a consumer, I want to query metadata in order to find recordings
- As a fieldworker, I want to keep input options as flexible as possible to be able to document musical performances precisely
- As a MIR expert, I want to retrieve consistent metadata for implementation in my algorithms
- As a scholar, I want to gain metadata on context of musical performances to involve them in my research.

To make metadata of archive assets consistent and comparable a standard needs to be defined, usually within a so-called *metadata schema*. Bradley [3] remarks, that laying out a custom metadata schema is not a trivial task and involves close cooperation between audio technicians, programmers and experts of the targeted domain. Further on he identifies the future value of a thorough layout process:

> Metadata is like interest–it accrues over time. If thorough, consistent metadata has been created, it is possible to predict this asset being used in an almost infinite number of new ways to meet the needs of many types of user, for multi-versioning, and for data mining [3, p. 14].

According to Bradley [3], metadata can be categorized into descriptive, structural and administrative sets of values. Descriptive metadata describes the content of the recording and is used in discovery and identification. Structural metadata includes structural information on the internal organization of an object, for example related graphics or temporal annotations. Administrative metadata holds information important for the management of an object, for example dates related to changes of a dataset, rights and licensing information or file formats. Metadata of three categories will usually be present in metadata schemas [3].

Creating an own metadata schema, thoughts need to be given not only to naming and structuring metadata attributes, but also to the values stored within the metadata schema. Types of values can range from simple numeric and textual to more complex representations, for example as a product of MIR feature extraction. Metadata derived from MIR and musical audio mining processes will among others include serializable data types like vectors and matrices for temporal and spatial representations.

A challenge was identified in proving standardization of data values throughout the archive. A trade-off between flexibility and standardization will be expected. Providing free text fields for descriptive metadata would be the most flexible solution, but it will most likely result in inconsistent or ambiguous descriptions. Existing archive's corpora already suffer from inaccurate metadata as "fuzzy matching shows

that fields are often similar but not exactly the same" [28, p. 109]. One solution can be validating data input for specific metadata values. Formalized data types such as time and date values can be validated well by nature. For textual data types, a solution is the use of controlled vocabularies. Controlled vocabularies limit the options, that a user can select from when creating or editing an archive asset (for example a list of predefined instruments to choose from instead of a free-text field).

The use of controlled vocabularies offers the chance to add semantic value to sets of metadata as well. Relations such as taxonomies or many-to-many relationships can be modeled, language translations can be introduced or synonymous relationships can be expressed using thesauri (for example employing the *American Folklore Society Ethnographic Thesaurus*[9]). Such relations can be of use in data mining processes or can support search queries by users. A use case for taxonomies in an ethnomusicological archive could be utilizing von Hornbostel's and Sachs' taxonomy for musical instruments [32]. Implementing this taxonomy, the archive could for example be search-queried by 'chordophones' returning musics instrumented by 'dutar', since a dutar is classified being a type of string instrument.

Gough and Han [9] point out, that it should be a priority for archives to find out what their individual requirements according metadata in terms of access and preservation are—the standards can be met later "as an end product generated by a tool included in the document" [9, p. 45]. They expect, that custom sets of metadata can usually be mapped or transformed to industry standards to enable interoperability.

The initial list of user stories can be extended as follows:

- As an archivist, I want to limit input options to keep metadata consistent
- As an archivist, I want to provide mapping internal to external metadata schemas in order to enable interoperability.

Mapping an archives internal metadata schema to Dublin Core (DC)[10] is a basic requirement that needs to be satisfied for OAIPMH compatibility. When working with aggregators such as cultural heritage platforms mentioned in Sect. 3.2, foreign metadata schemas need to be satisfied. OAIPMH is capable of providing multiple metadata schemas side by side. This enables metadata harvesters to request a specific schema and helps sound archives to integrate with multiple harvesters. Common standards required by harvesters include Metadata Encoding and Transmission Standard (METS),[11] Metadata Object Description Schema (MODS)[12] or Europeana Data Model (EDM).[13] Since all mentioned standards are based on XML technology, so-called *extension schemas* can be introduced to cover specific requirements of an archive [3].

[9] http://id.loc.gov/vocabulary/ethnographicTerms.html.

[10] http://dublincore.org.

[11] http://www.loc.gov/standards/mets.

[12] http://www.loc.govstandards/mods.

[13] https://pro.europeana.eu/resources/standardization-tools/edm-documentation.

3.4 Implementation of MIR

Notated scores and transcriptions build the foundation for a large amount of existing research in all branches of musicology. Since a lot of music doesn't exist in form of scores, scholars of ethnomusicology from the discipline's advent have involved analysis of audio recordings into their research by transcribing them by ear. Since the last two decades, sound archive's music assets can be analyzed in various ways using computational methods, which is of growing interest to ethnomusicologists [31]. User stories related to MIR and sound archives brought up by stakeholders included:

- As a fieldworker, I want to compare my private recordings to archive's public contents to identify similarities
- As a MIR expert, I want identify similarities of audio recordings to help people to retrieve music
- As a scholar, I want to be able to export results of MIR feature extraction for usage in my research
- As a scholar, I want to take advantage of automatic transcriptions to support or even avoid manual transcription.

Marsden [16] identifies three main capabilities of the computer in musicological research. (1) the computer is able to be neutral and impersonal, (2) it can deal with large quantities of music, and (3) the computer can reveal what cannot be heard by the human ear. All three capabilities can be seen as crucial to the discipline of MIR, since it deals with large amounts of music and relies on empirical analysis of complex data structures that could hardly be executed by humans.

One core activity of MIR is the extraction of musical features to adequately describe the sound signal by numeric feature vectors or textual values that can be processed further on by computers [17]. Three main feature classes of sound recordings can be identified: conceptual metadata, high-level descriptions of musical content and low-level properties of music [4].

Conceptual metadata is usually represented as text and therefore can be searched using textual queries. Conceptual metadata can be subdivided into factual and cultural metadata. Factual metadata describes objective truth about a piece of music, for example name of artist, year of publication, track title or duration. Cultural metadata includes subjective descriptions of music such as mood, emotion, genre or style. Conceptual metadata can be collected or elicited by humans or can be derived from feature extraction algorithms [4]. Popular examples are genre or mood classifiers.

High-level descriptions employ musical concepts like melody, rhythm or harmony to represent the content of audio files. High-level descriptions link with intuitive or expert knowledge of how a song is constructed and thus can be either derived from expert analysis or can be extracted from digital audio using MIR algorithms [4]. As an example, a high-level feature extraction could be applied to detect the serial pitches of a single line melody.

Low-level features are representations of digital measurements of audio signals and usually can only be extracted and processed in a computational way. Low-level feature extraction can be segmented in three different ways: (1) frame based segmentations (sampling at 10–1000 ms intervals), (2) beat-synchronous segmentations (features are extracted on beat) and (3) statistical measures that produce probability distributions from features. A common downstream approach is to extract metadata or high-level representations from low-level features [4].

Having access to conceptual metadata, high- and low-level features, music collections can be structured and queried in novel ways. Leman [14] described these tasks as *Musical Audio Mining*:

> Musical audio mining can be defined as data mining on musical audio. It will allow users to search and retrieve music, not only by means of text queries (such as title, composer, text song, conductor, orchestra), but also by means of content-based musical queries, such as query-by-humming/singing/playing, by specification of a list of musical variables (such as 'happy', 'energetic', etc.), or by means of given sound excerpts, or any combination of audio and text [14, p. 440].

Musical audio mining can be seen as a branch of *Data Mining*. Like in data mining, classification, similarity detection or categorization of large datasets will utilize well known statistical methods or machine learning concepts like neural networks or self-organizing maps (SOMs) [5]. These concepts help to abstract from high dimensional feature spaces to less complex data representations for example conceptual metadata or can be used to detect coherencies between multiple recordings. Musical audio mining can help to draw conclusions about the musical content and the relation of different musical contents to each other.

Historically, MIR and musical audio mining are focused on Western popular and classical music. Ethnomusicological sound archives raise new tasks for the discipline. It might be especially of interest to search for distinctive musical features that are specific for a musical tradition instead of only searching for similarities of sounds. This practice could lead to the identification of cultural or geographic affiliations of recordings or musical traditions. Additionally, already existing conceptional metadata can be used by musical audio mining algorithms, since assets in ethnomusicological sound archives are usually well documented [24].

As shown in recent publications, employing MIR technology, ethnomusicological sound archives can already profit from the three unique capabilities of computers that Marsden [16] mentioned.

(1) Since the computer is not a trained musicologist or music expert, musics from different cultural backgrounds—including distinct musical concepts like tonal systems or instruments—can be transcribed, empirically analyzed and related *as is* without any prejudices. As an example, cross-cultural mood classifiers could be identified, as approached by Yang and Hu [35] for a set of English and Chinese songs.

(2) Manual transcription of musical performances is a slow and demanding expert task, thus limiting the number and scale of investigations [33]. Computers are capable of automating the transcription process of high-level features and even transcribe conceptional metadata like for example instrumentation in ethnomusicological recordings [7]. Automatic transcription of ethnic and folk music recordings of

specific genres was recently explored by Gong et al. [8] as well as Holzapfel and Benetos [11]. Results from feature extraction can be fed back into an archive's user interface. Archive projects already utilizing such capabilities include *Telemeta*[14] and *Dunya*.[15,16] Telemeta's player offers visualizations of low- and high-level features such as pitch or spectrogram extracted by MIR framework *TimeSide*.[17] The user interface of Dunya provides visualizations of features like rhythm and pitch as well as real-time score and lyric following [23].

Another approach to discover similarities, distinctions or coherencies between recordings is analyzing large-scale music collections as done by the *Digital Music Lab*[18] (DML) project. A software is being developed that is capable of extracting various features from audio and offers an interface that that helps analyzing and comparing results on a collection level [1].

A practical use case for a an interface utilizing feature extraction of *tempo* was brought up at a concert and discussion with musicians from Xinjiang that took place at the *Institute for Systematic Musicology* in 2016. Among the musicians were some members of the ethnic group of *Uyghurs*, performing songs in a traditional music style *muqam*. The musicians were asked, if they can confirm a trend over the last years, to slowing down the musical performance incorporating more melodic details. If a sound archive had access to multiple recordings of muqam over a time period and tempo would be extracted automatically, a graphical representation or even reviewing a list of all recordings could provide a quick answer.

(3) Blaß [2] showed, that low-level features like for example timbre can be extracted and have the potential to represent high-level features—in this case rhythm—in novel ways. Extraction and post-processing of low-level features could potentially result in representations of musical universals that are closer to the human perception than well-known representations. Statistical- as well as machine learning algorithms can be used to identify and visualize coherencies between music cultures from different parts of the world—for example in form of topographical maps as examined by Juhász [12].

Additionally, MIR can play an important role in automating archiving tasks and solving common archiving problems. Six et al. [28] presented a practical solution to detect duplicates in a collection of African field recordings. They were able to show that acoustic fingerprinting algorithms are capable of dealing with differences in speed (duplicates might have been copied or digitized with equipment that was not well calibrated) and that metadata of duplicates might vary. Merging of duplicate's metadata can lead to better archive quality [28].

---

[14] http://telemeta.org, developed as a cooperation between *Center for Research in Ethnomusicology* at the *Université Paris Ouest* and the *Musée de l'Homme*.

[15] http://dunya.compmusic.upf.edu, developed as a part of *CompMusic* at *Universitat Pompeu Fabra*.

[16] http://compmusic.upf.edu.

[17] https://github.com/Parisson/TimeSide.

[18] http://dml.city.ac.uk, developed in cooperation between *City University London*, *Queen Mary University of London*, *University College London* and the *British Library*.

The list of user stories addressing implementation of MIR in sound archives can be extended as follows:

- As an archivist, I want to utilize MIR to discover and clean up duplicates as well as enrich metadata
- As an scholar, I want to compare MIR results for sets of recordings to discover similarities and distinctions.

Stakeholders of groups *MIR experts* and *developers* brought up the following functional remarks related to the introduction of MIR to sound archives.

First, it was considered important to support different implementations of feature extraction algorithms, since an archive might implement multiple algorithms that are capable of extracting the same feature. For example different implementations to extract single-line melodies from an audio recording should all be made available side by side. Algorithms might as well be optimized or changed over time, so once results of a feature extraction would be published, users or API implementations must be able to toggle between different versions, to provide stability of results. It was mentioned that the *AcousticBrainz*[19] project already developed an internal standard including versioning for utilized software libraries.

The implementation of various MIR algorithms might be supported by so called *MIR toolboxes*. A lot of research has been done in the area of feature extraction and many algorithms already exist as implementations. A recent evaluation of toolboxes was published by Moffat et al. [18]. Since existing algorithms might be written in a variety of programming languages, a multi-language support was considered mandatory.

Since feature extraction or audio mining algorithms are often time and resource consuming, it might not be feasible to apply them in real-time. It was considered important, that new recordings being added to the archive are analyzed asynchronously in background processes. To optimize overall performance on large collections it was considered desirable to implement parallelization for such processes.

Furthermore, datasets and visualizations resulting from MIR algorithms should be exportable to common data exchange formats like lossless bitmaps, vector graphic formats or non-proprietary text formats such as Comma-separated Values (CSV). This will support further processing outside of the archive and will enable scholars to use results in publications.

3.5 CoMSAr Project

Based on requirements artefacts derived from user interviews and literature, implementation of CoMSAr was executed.

[19] https://acousticbrainz.org.

3.5.1 Technology and Architecture

For CoMSAr, preferably popular open source technology with high market share was involved. It was expected that this would ensure long term support and good documentation of technologies. CoMSAr's backbone is provided by the popular *LAMP Stack* (Linux operating system, Apache HTTP Server,[20] MySQL relational database management system[21] and PHP programming language[22]). On top of PHP, *Zend Framework*[23] was implemented. Zend is an object-oriented, open source web application framework written in PHP and backed by a vibrant open source community. Zend offers developers, among other components, a high performance Model-view-controller (MVC) implementation and a robust and secure database abstraction layer.

To enable *Test-driven Development* (TDD) and to support robustness of the system in agile development processes, the *Codeception*[24] framework for PHP was integrated. Codeception offers writing and executing user acceptance tests. Codeception emulates a web browser and offers actions like navigating to specific URLs, populating form fields, clicking interface elements and verifying if specific HTML elements or contents exist. Tests were written for many use cases including creating and updating new archive records.

Architecture of CoMSAr can be visualized as a client-server relationship, whereas client and server can be seen as a hierarchy of technologies involved (Fig. 1). A client could be a user's computer, that renders content served employing the HTTP protocol. In this case, returned content types could be HTML, XML, Cascading Style Sheets (CSS), media files or JavaScript data. A client could be a web service accessing content through an OAIPMH API as well. Figure 1 shows web services X, Y and Z as placeholders for an unknown quantity of foreign servers. Those could be for example other university's archives that query the database or cultural heritage platforms harvesting metadata. In this case the the server will return OAIPMH compliant XML data.

As mentioned in Sect. 3.1, the front-end needs to adapt to different screen sizes, this could be achieved by implementing concepts of *Responsive Web Design* (RWD). The term RWD was coined by Ethan [15]. RWD makes use of CSS media queries, a grid system and flexible media and images. Media queries are used serve distinct CSS rules based on the user's browser width (for example to hide elements on small screens). A grid system is used to transform multi-column layouts to single-column layouts on small mobile devices. Column widths are usually set on a percentile basis to relate and adapt to the browser window's width. Images are usually set to fill up 100% of the column width and preserve the aspect ratio, in that way they adapt to the screen size [15].

[20] http://apache.org.
[21] https://www.mysql.com.
[22] http://www.php.net.
[23] http://www.zend.com.
[24] http://codeception.com.

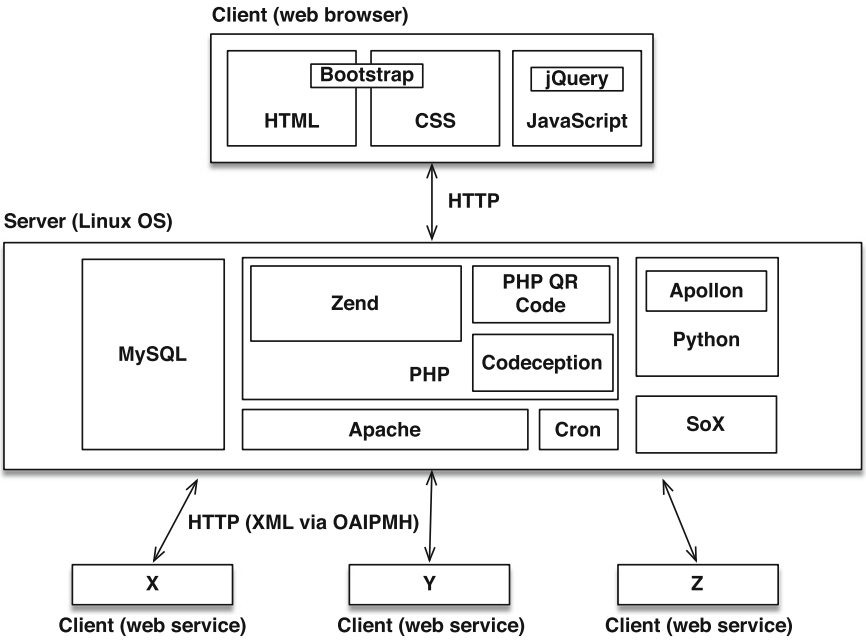


Fig. 1 Client-server relationship of CoMSAr, revealing the hierarchy of technologies involved

All CSS rules were created with the *mobile first* approach in mind. The term *mobile first* was coined by Wroblewski [34]. Rules applying to all screen sizes are written in a main CSS file. In a second CSS file specific rules for larger screen sizes are overridden, if necessary. Three main break points for layout changes are introduced utilizing media queries. Extra-small mobile devices with a maximum screen size of 768 px, small devices, tablets and laptops with screen sizes smaller than 1200 px and large screens of desktop computers with screen sizes larger than 1200 px (Fig. 2).

Since for CoMSAr the *Bootstrap*[25] framework was implemented, most HTML elements are ready for RWD by default. Bootstrap's grid system is utilized to create one- and two-column layouts, whereas the two-column layout transforms into a one-column layout on smaller screen sizes. For the attribute listings of the *recording detail views* the HTML elements of *description lists* (`dl`, `dt`, `dd`) were used, since description lists can be formatted dynamically either in a horizontal order (desktop) or in a vertical order (tablet and mobile). The table of the *explore view* can be scrolled vertically on tablet and mobile devices.

An expandable mobile navigation was implemented to save valuable screen space on mobile devices.

[25] https://getbootstrap.com.

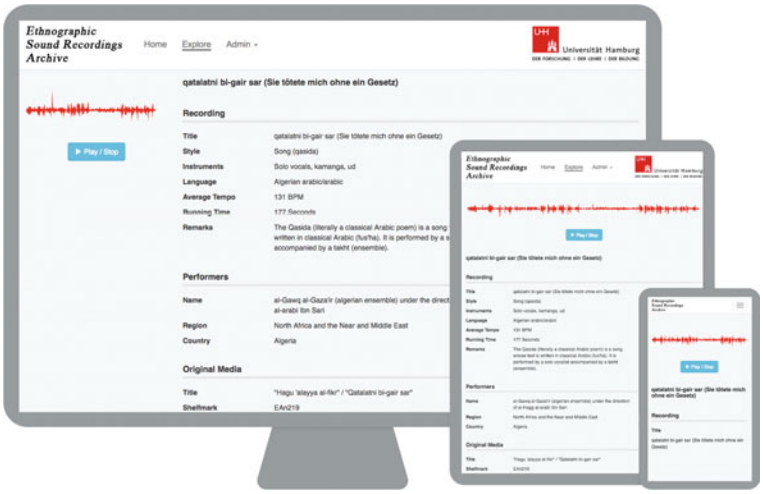


Fig. 2 CoMSAr implementation for E.S.R.A. adapting to three different device screens (*desktop*, *tablet* and *mobile*) by implementing RWD

3.5.2 Features

Besides table stake features like a user and permission management, registration, full-text search interface, list views, an audio player and views to create and edit archive contents, the following features were implemented as of now.

To support MIR feature extraction algorithms, a background process scheduler was developed. Since implementations of MIR algorithms will not necessarily be written in the same programming language as the database application, the process scheduler is setup to run scripts on the command line, passing parameters, retrieving the processed results and storing them in the database. Feature extraction can be run asynchronously and parallelized for multiple sound files to optimize the use of resources.

CoMSAr implements *Apollon*[26] framework that is capable of extracting and classifying timbre states of sound files, calculate the transition probability of all states employing *Hidden Markov Models* (HMM) and using transition probabilities of multiple sound files to train Kohonen maps. The configuration of a Kohonen map can be saved once trained and new sound files can be located on the two-dimensional map (Fig. 3).

To encourage scholars to use archive audio files as primary resources, two ways to support easy referencing were implemented. First, on the every recording view, a *BibTeX*[27] representation of the metadata is offered for copy and pasting into users BibTex library files. To offer a second, more convenient option, recording's attributes

[26] Developed by Michael Blaß.

[27] http://www.bibtex.org.

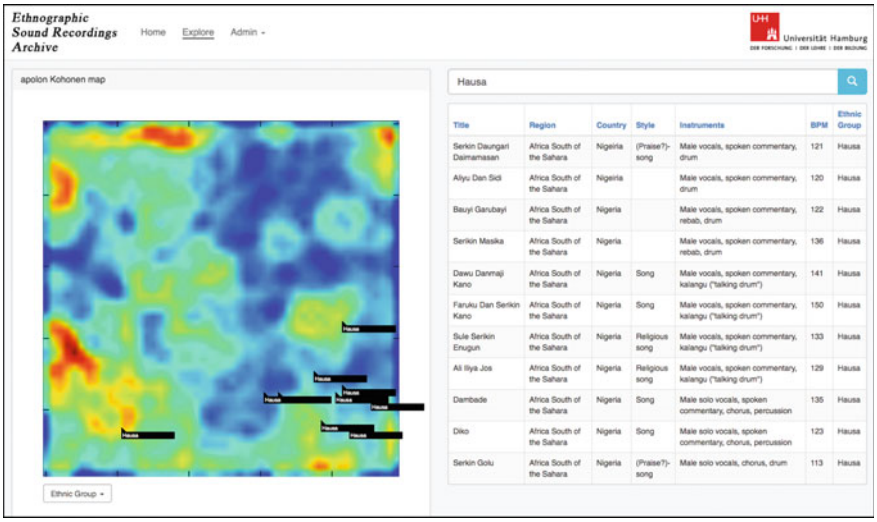


Fig. 3 The explore view showing search results for search term *Hausa*

are implemented in the HTML markup of each recording view. The HTML header includes *meta tags*, employing the Dublin Core standard. Citation management software like *Zotero*[28] often offers plug-ins for common web browsers that interpret the markup and, if annotated according to specific standards, can transform web site's content directly into bibliographic objects. In this way recordings archived in CoMSAr can be transformed into bibliographic objects by the click of a button.

To support easy referencing and playback of archive sound files in public spaces like for example museums, a *QR code* for each recording is created, can be downloaded and used for printing. QR codes can be scanned by mobile devices and will redirect users to recordings when scanned on a mobile device.

To enable CoMSAr to contribute to cultural heritage platforms and to allow other archives to access CoMSAr through an API, OAIPMH was implemented. The archive's internal metadata schema is mapped to Dublin Core schema, to provide basic OAIPMH compatibility.

4 Conclusions and Future Work

This chapter summarizes selected requirements artefacts discovered during development of CoMSAr. Results from user and expert interviews as well as literature research are taken into account. It can be concluded that open access on the WWW is a key to increase an archive's productivity. The availability of recordings online

[28] https://www.zotero.org.

will support cultural transmission processes and open the field for academic uses among others. A public API will enable archives to share contents with partner organizations, support large-scale research projects and broaden an archive's audience by disseminating contents on cultural heritage platforms. An elaborate metadata concept ensures consistency of data, will help to add semantic value and will be crucial to providing interoperability.

Between MIR and ethnomusicological sound archives a fruitful relationship can be established, since sound archives possess valuable recordings that are needed to create and optimize MIR algorithms beyond the scope of Western music. Feeding back results from MIR into archive's metadata can support for example building novel user interfaces to reveal coherencies between recordings, can provide automatic transcriptions and enriched metadata, or can help with archive maintenance tasks.

Requirements discovered were used to create an implementation of CoMSAr for the *Ethnographic Sound Recordings Archive*. The following future tasks are yet to be approached:

1. Define a metadata extension schema that meets future requirements of ethnomusicologists as well as of MIR experts and enables interoperability with other CoMSAr instances. The extension schema should extend Dublin Core schema in OAIPMH transactions.
2. Introduce controlled vocabularies and taxonomies for metadata attributes like *instrumentation* or *location of recording*. For *location of recording* the the biggest challenge is handling historical data that refers for example to countries that don't exist anymore. For *instrumentation*, an implementation of Hornbostel's and Sachs' taxonomy seems promising. But a review of existing datasets revealed that a dynamic way of extending taxonomies while entering data is needed.
3. Allow translations of metadata values. A precise documentation of the original ethnographic notes is important but to help computational analysis, mapping to more generic terms is often necessary. A way to store both options needs to be implemented.
4. Since some ethnographic notes include fuzzy data or some data can be only determined to a certain extent, a way to store fuzzy data needs to be introduced. This can be especially of use for date statements like for example '*before 1908*' or '*around 1943*' that could be found in existing datasets.
5. Graphical representations such as Kohonen maps or raw data sets resulting from MIR algorithms should be exportable by users.
6. Support multiple video or image attachments for recordings. In cases where only video exists, extract audio track from video.
7. Since field recording nowadays often utilizes multi-track recording devices, allow multi-track audio upload. Multiple audio files should be storable as a supplement and should be playable by users.

References

1. Abdallah S, Benetos E, Gold N, Hargreaves S, Weyde T, Wolff D (2016) Digital music lab: a framework for analysing big music data. In: Signal processing conference (EUSIPCO 2016), Budapest, Hungary, pp 1118–1122
2. Blaß M (2013) Timbre-based drum pattern classification using hidden Markov models. In: The European conference on machine learning and principles and practice of knowledge discovery in databases (ECMLPKDD 2013), Czech Republic, Prague, pp 11–14
3. Bradley K (ed) (2009) Guidelines on the production and preservation of digital audio objects, 2nd edn. In: IASA, Auckland Park
4. Casey M, Veltkamp R, Goto M, Leman M, Rhodes C, Slaney M (2008) Content-based music information retrieval: current directions and future challenges. Proc IEEE 96(4):668–696
5. Clifton C (2009) Data mining. https://www.britannica.com/technology/data-mining
6. Feld S (2002) Sound recording as cultural advocacy: a brief case history from Bosavi, Papua New Guinea. In: Berlin G, Simon A (eds) Music archiving in the world: papers presented at the conference on the occasion of the 100th anniversary of the Berlin Phonogramm-Archiv. VWB-Verlag, Berlin, Germany
7. Fourer D, Rouas J-L, Hanna P, Robine M (2014) Automatic timbre classification of ethnomusicological audio recordings. In: International society for music information retrieval conference (ISMIR 2014), Taipei, Taiwan, pp 295–300
8. Gong R, Yang Y, Serra X (2016) Pitch contour segmentation for computer-aided Jingju singing training. In: 13th sound and music computing conference (SMC 2016), Hamburg, Germany, pp 172–178
9. Gough J, Han M-JK (2014) Creating metadata best practices for digital audiovisual resources. IASA J 43:37–47
10. Gray J (2002) Performers, recordists, and audiences: archival responsibilities and responsiveness. In: Berlin G, Simon A (eds) Music archiving in the world: papers presented at the conference on the occasion of the 100th anniversary of the Berlin Phonogramm-Archiv. VWB-Verlag, Berlin, Germany, pp 48–53
11. Holzapfel A, Benetos E (2016) The Sousta corpus: beat-informed automatic transcription of traditional dance tunes. In: 17th international society for music information retrieval conference (ISMIR 2016), New York City, USA, pp 531–537
12. Juhász Z (2011) Low dimensional visualization of folk music systems using the self organizing cloud. In: 12th international society for music information retrieval conference (ISMIR 2011), Miami, USA, pp 299–304
13. Lehmann K-D (2002) Greeting dddress. In: Berlin G, Simon A (eds) Music archiving in the world: papers presented at the conference on the occasion of the 100th anniversary of the Berlin Phonogramm-Archiv. VWB-Verlag, Berlin, Germany, p 13
14. Leman M (2002) Musical audio mining. In: Meij J (ed) Dealing with the data flood: mining data, text and multimedia, SST, Rotterdam, pp 440–456
15. Marcotte E (2011) Responsive web design. A book apart, 2nd edn. New York
16. Marsden A (2009) "What was the question?": music analysis and the computer. In: Gibson L, Crawford T (eds) Modern methods for musicology. Ashgate, Farnham, England, pp 137–147
17. Mayer R, Frank J, Rauber A (2009) Analytic comparison of audio feature sets using self-organising maps. In: Proceedings of the workshop on exploring musical information spaces (WEMIS 2009), Corfu, Greece, pp 62–67
18. Moffat D, Ronan D, Reiss JD (2015) An evaluation of audio feature extraction toolboxes. In: 18th international conference on digital audio effects (DAFx-15), Trondheim, Norway, pp 71–77
19. Nelson-Strauss B, Brylawski S, Gevinson A, National Recording Preservation Board (U.S.) (eds) (2013) The library of congress national recording preservation plan. Number No. 156 in CLIR publication. Council on Library and Information Resources: Library of Congress, Washington, D.C

20. Panteli M, Benetos E, Dixon S (2018) A review of manual and computational approaches for the study of world music corpora. J New Music Res 47(2):176–189
21. Pohl K (2010) Requirements engineering–fundamentals, principles, and techniques. Springer, Heidelberg
22. Porter A, Bogdanov D, Kaye R, Tsukanov R, Serra X (2015) Acousticbrainz: a community platform for gathering music information obtained from audio. In: Proceedings of the 2015 international society for music information retrieval (ISMIR 2015), Málaga, Spain, pp 786–792
23. Porter A, Sordo M, Serra X (2013) Dunya: a system for browsing audio music collections exploiting cultural context. In: Proceedings of 14th international society for music information retrieval conference (ISMIR 2013), Curitiba, Brazil
24. Proutskova P (2007) Musical memory of the world–data infrastructure in ethnomusicological archives. In: Proceedings of the 2007 international society for music information retrieval (ISMIR 2007), Vienna, Austria, pp 161–162
25. Seeger A (1986) The role of sound-archiving in ethnomusicology today. Ethnomusicology 30
26. Seeger A (2002) Archives as part of community traditions. In: Berlin G, Simon A (eds) Music archiving in the world: papers presented at the conference on the occasion of the 100th anniversary of the Berlin Phonogramm-Archiv. VWB-Verlag, Berlin, pp 41–47
27. Simon A (2002) The Berlin Phonogramm-Archiv and the preservation of unwritten music. In: Berlin G, Simon A (eds) Music archiving in the world: papers presented at the conference on the occasion of the 100th anniversary of the Berlin Phonogramm-Archiv. VWB-Verlag, Berlin, Germany, pp 32–36
28. Six J, Bressan F, Leman M (2018) Applications of duplicate detection in music archives: from metadata comparison to storage optimisation. In: Serra G, Tasso C (eds) Digital libraries and multimedia archives, vol 806. Springer International Publishing, Cham, pp 101–113
29. Thedens H-H (2002) Local archives as a resource for the living folk music tradition: recent developments in Norway. In: Berlin G, Simon A (eds) Music archiving in the world: papers presented at the conference on the occasion of the 100th anniversary of the Berlin Phonogramm-Archiv. VWB-Verlag, Berlin, Germany, pp 70–78
30. Tzanetakis G (2014) Computational ethnomusicology: a music information retrieval perspective. In: International computer music conference proceedings (ICMC 2014), Greece, Athens, pp 69–74
31. Tzanetakis G, Ajay K, Schloss WA, Wright M (2007) Computational ethnomusicology. J Interdiscip Music Stud 1(2):1–24
32. von Hornbostel E, Sachs C (1961) The classification of musical instruments. Galpin Soc J 3(25):3–29
33. Weyde T, Cottrell S, Dykes J, Benetos E, Wolff D, Tidhar D, Kachkaev A, Plumbley M, Dixon S, Barthet M, others (2014) Big data for musicology. In: Proceedings of the 1st international workshop on digital libraries for musicology (DLfM 2014), London, United Kingdom, pp 85–87
34. Wroblewski L (2011) Mobile first. A book apart, New York
35. Yang Y-H, Hu X (2012) Cross-cultural music mood classification: a comparison on English and Chinese songs. In: Proceedings of the 2012 international society for music information retrieval (ISMIR 2012), Porto, Portugal, pp 19–24

Part IV
Physical Modeling and Measurements

Laser-Based Interferometric Techniques for the Study of Musical Instruments

Efthimios Bakarezos, Yannis Orphanos, Evaggelos Kaselouris, Vasilios Dimitriou, Michael Tatarakis and Nektarios A. Papadogiannis

Abstract Laser-based interferometry techniques for the study of musical instruments are discussed in this chapter. The presented work demonstrates the advantages that the laser-based optical techniques provide and is mainly focused on the capabilities of the Electronic Speckle Pattern Interferometry (ESPI). The mathematical description of the time-average ESPI, the reading of the contour lines and the overcoming of the limitations of amplitude and phase vibration are analyzed. Furthermore, four representative studies using the ESPI for a Cretan Lyra, a Bendir, a classical Guitar and the ancient Greek lyra Chelys, are presented and demonstrate the capabilities of the method.

1 Introduction

The study of vibrations and the development of vibration analysis and measurement techniques are important for a great number of applications in many and various fields, such as the automotive and aircraft industry (e.g. car brakes, car parts, turbine blades), material science and technology (e.g. defect detection, impact damages), audio industry (e.g. loudspeaker quality control), musical acoustics (e.g. musical instruments vibration analysis), medicine (eardrum diagnostics), electronics (hard disk heads), etc.

Laser-based optical techniques, such as holographic interferometry [14], laser Doppler vibrometry [6], laser Doppler velocimetry [25], offer a number of unique advantages over traditional mechanical techniques, such as impact hammer/impulse

E. Bakarezos · Y. Orphanos · N. A. Papadogiannis (✉)
Department of Music Technology & Acoustics Engineering, School of Applied Sciences, TEI of Crete, Heraklion, Greece
e-mail: npapadogiannis@staff.teicrete.gr

E. Bakarezos · Y. Orphanos · E. Kaselouris · V. Dimitriou (✉) · M. Tatarakis · N. A. Papadogiannis
School of Applied Sciences, Centre for Plasma Physics and Lasers, TEI of Crete, Heraklion, Greece
e-mail: dimvasi@chania.teicrete.gr

R. Bader (ed.), *Computational Phonogram Archiving*, Current Research in Systematic Musicology 5, https://doi.org/10.1007/978-3-030-02695-0_12

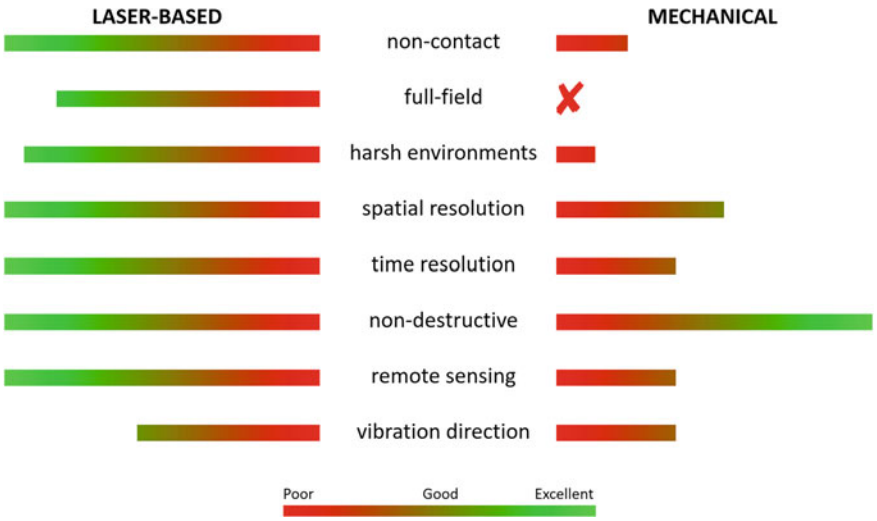


Fig. 1 Laser-based versus mechanical techniques

response analysis, use of piezoelectric elements and microphone arrays. While in conventional photography only the light intensity is recorded, in holography the phase is recorded as well. Compared to point or scanning laser-based techniques like Doppler vibrometry, holographic or holographic interferometry techniques have the advantage of full-field imaging. A comparison of the laser versus the mechanical based techniques is schematically presented in Fig. 1.

2 Holographic and Speckle Interferometry

Holographic interferometry is a combination of interferometry and holography [5, 14, 17]. The interferometric comparison occurs for two, or more, wave fields, of which at least one is holographically reconstructed. The main difference between classical holography and holographic interferometry is that the object beam is compared with itself: typically, the object stays in its place, and if between two successive recordings the object is somehow deformed, the relative phases of the two light fields will alter and it is possible to observe the interference.

With the advent of technology, CCD/CMOS cameras are used as the recording medium, hence the terms "TV-holography" or "electronic holography". In Electronic Speckle Pattern Interferometry (ESPI) one of the interfering fields is the speckle pattern that results from the reflection of coherent light off an optically rough surface [17, 20, 26]. The speckle effect is a result of the interference of many waves of the same frequency, having different phases and amplitudes. These waves superimpose to give a resultant wave whose amplitude and therefore intensity, varies randomly.

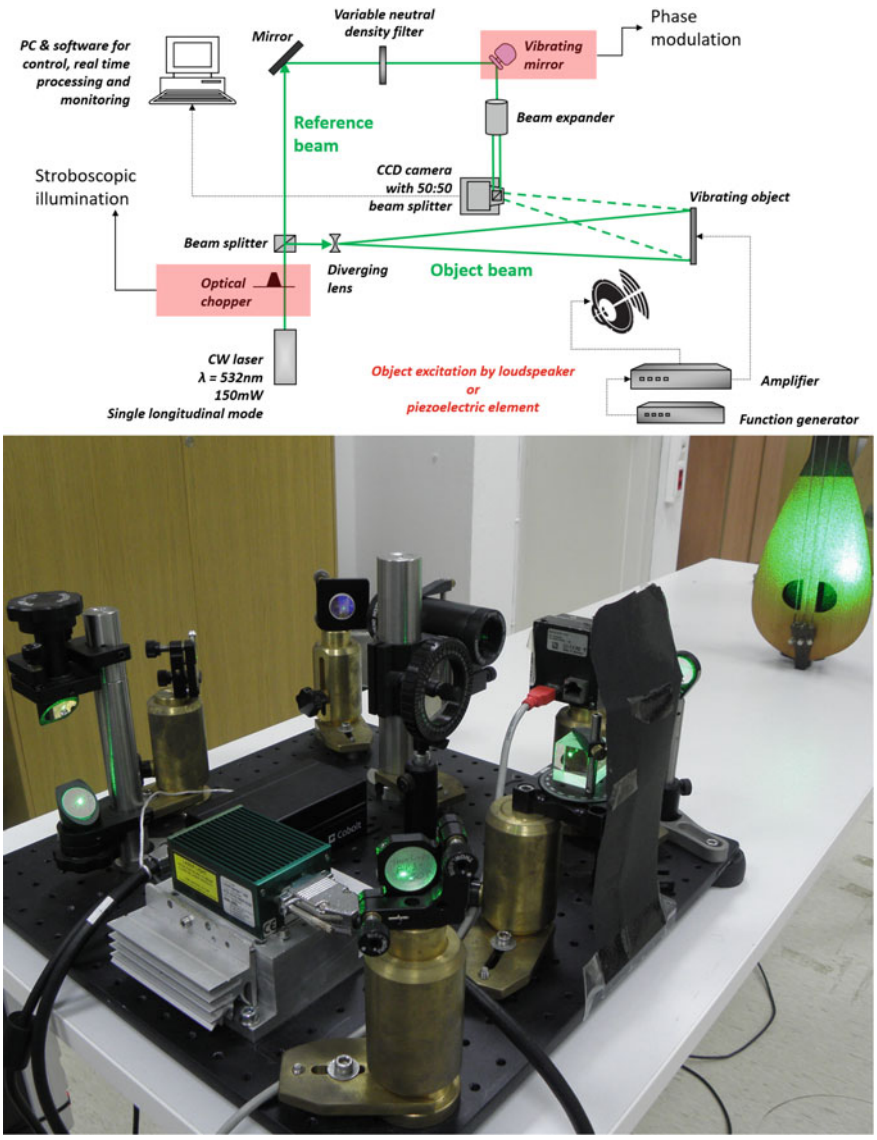


Fig. 2 Time-average ESPI—experimental set-up (top) and the portable ESPI system of the lab (bottom)

In Fig. 2 the experimental Set-up diagram of the time-average ESPI (Top) and its portable prototype laboratory implementation (Bottom) are presented.

The mathematical description of the time-average ESPI, the reading of the contour lines and the overcoming of the limitations of amplitude and phase vibration are further analyzed at the following subsections.

2.1 Time-Average ESPI Mathematical Description

For the out of plane vibrating surface, on the z axis direction, it holds that:

$$w(x, y, t) = A(x, y)\cos[2\pi ft + \varphi_0(x, y)] \tag{1}$$

where:

$A(x, y)$ surface vibration amplitude spatial distribution
f frequency of the vibrating surface
$\varphi_0(x, y)$ surface vibration phase spatial distribution beams.

With regard to the intensity distribution of the first recorded image it holds that:

$$I_1 = \frac{1}{\tau}\int_0^{\tau}\left\{I_{OBJ} + I_{REF} + 2\sqrt{I_{OBJ}I_{REF}} \cdot \cos\left[\varphi + \frac{2\pi}{\lambda}(1 + cos\theta)\right.\right.$$
$$\left.\left. A\cos(\omega t + \varphi_0)\right]\right\}dt \tag{2}$$

where:

τ CCD camera exposure time
I_{OBJ} object beam
I_{REF} reference beam
θ angle between object and reference beams
φ phase difference between object and reference beams
λ laser wavelength (532 nm)
ω angular frequency.

For $\Gamma = 4\pi/\lambda$ ($\theta \approx 0°$) and $\tau = 2\, m\pi/\omega$, where m is the mode number, it becomes [14, 18]:

$$I_1 = I_{OBJ} + I_{REF} + 2\sqrt{I_{OBJ}I_{REF}}|(\cos\varphi)J_0(\Gamma A)| \tag{3}$$

where J_O the solution of the Bessel differential equation.

The intensity distribution of the second recorded image (where $A \rightarrow A + \Delta A$) becomes [12]:

$$I_2 = I_{OBJ} + I_{REF} + 2\sqrt{I_{OBJ}I_{REF}}\left|(\cos\varphi)[1 - \frac{1}{4}\Gamma^2(\Delta A)^2]J_0(\Gamma A)\right| \tag{4}$$

After the subtraction of images ($I_2 - I_1$) the total intensity is:

$$I = I_2 - I_1 = \frac{\sqrt{I_{OBJ}I_{REF}}}{2}\left|(\cos\varphi)\Gamma^2(\Delta A)^2 J_0(\Gamma A)\right| \tag{5}$$

In this analysis the interference fringes are modified by a Bessel function of the first kind zero order and fringes of decreasing brightness are created.

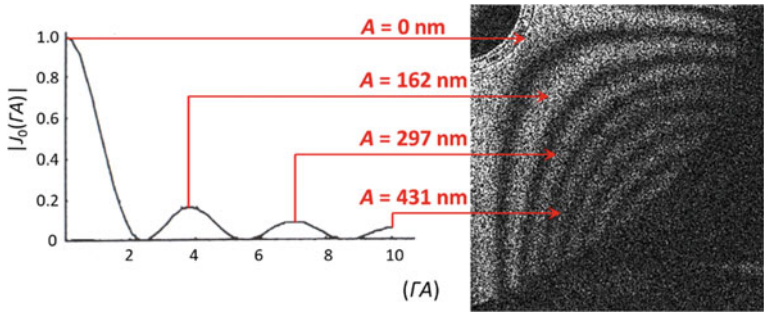


Fig. 3 Bessel function contour lines

2.2 Reading the Contour Lines of the Time-Average ESPI

Interference fringes are contour lines, i.e. lines of equal vibration amplitude. For $\lambda =$ 532 nm it holds:

$$\Gamma A = \frac{4\pi}{\lambda} A = 0, 3.830, 7.015, 10.175, \ldots (\textit{Bessel function argument values at local maxima}) \quad (6)$$

The brightest fringe corresponds to zero vibration amplitude (stationary regions).

As the vibration amplitude becomes larger the number of bright fringes increases as may be observed in Fig. 3.

2.3 Time-Average ESPI Limitations

The main limitations of the time-average ESPI is the amplitude and the phase, since for large vibration amplitudes (~ few tens of μm) bright fringes become indistinguishable and the surface vibration phase information is lost, i.e. term $\varphi_0(x, y)$ vanishes upon integration. These limitations may be resolved, as described in the two following paragraphs.

2.3.1 Resolving the Vibration Amplitude Limitations

To resolve the vibration amplitude limitations the stroboscopic illumination via the use of optical chopper to modulate the laser beam may be applied. The object is illuminated at two points of its vibration cycle. By this way, fringes of equal brightness ($\propto \cos^2$) are generated according to Eq. 7:

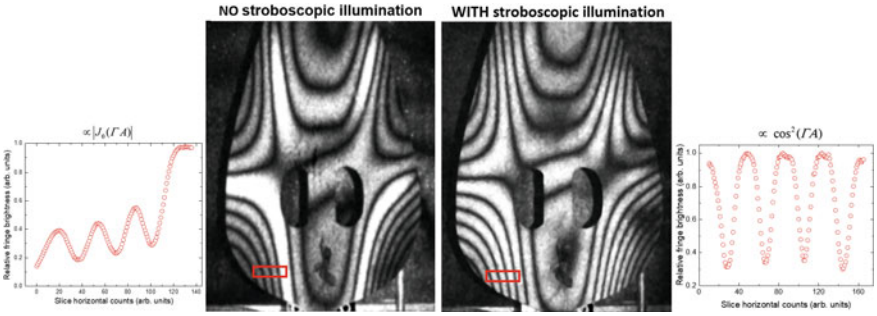


Fig. 4 Stagakis-type Cretan lyra. Resonance frequency 1106 Hz with and without stroboscopic illumination

$$I_{STROBOSCOPIC}(x, y) \propto I_0(\mathrm{x}, \mathrm{y}) \cos^2(\Gamma A) \tag{7}$$

In Fig. 4 are depicted ESPI vibrating results with and without stroboscopic illumination for a Cretan lyra.

2.3.2 Resolving Vibration Phase Limitations

In order to resolve the limitation of the vibration phase, Phase Modulation (PM) is used and more particularly a piezoelectric element is used to vibrate the reference beam mirror at the same frequency f as the vibrating object, with the ability to vary the phase.

The vibrating mirror displacement is:

$$u_{MIRROR}(x, y, t) = A_{MIRROR} \cos(2\pi f t + \varphi_{MIRROR}) \tag{8}$$

where:

A_{MIRROR} vibrating mirror amplitude
F vibrating mirror frequency = frequency of vibrating object
φ_{MIRROR} vibration phase of mirror.

$$I_{PHASEMODULATED}(x, y) \propto I_0(\mathrm{x}, \mathrm{y}) \Big\{ J_0^2 (\Gamma [A_{OBJECT}^2 + A_{MIRROR}^2 - 2A_{OBJECT} A_{MIRROR} \cos(\varphi_{OBJECT} - \varphi_{MIRROR})]^{\frac{1}{2}}) \Big\} \tag{9}$$

For the experimental implementation, the mirror vibrates at an amplitude of 0.4 μm with a maximum offset of 3 μm. An adjustment offset is used to achieve the desired PM. In Fig. 5 are depicted ESPI vibration results using mirror offset for a Cretan lyra.

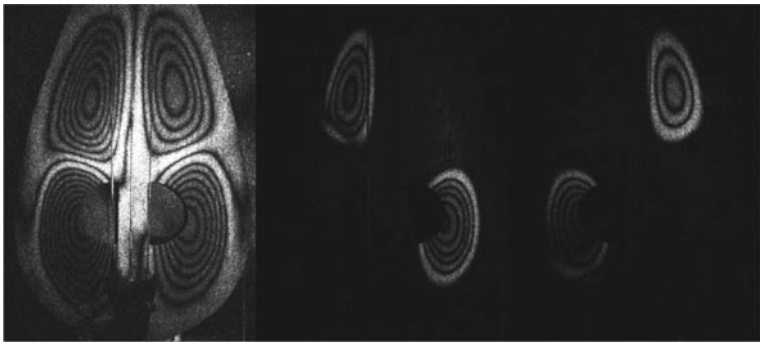


Fig. 5 Stagakis-type Cretan lyra. (Left) Resonance frequency 1290 Hz, (center) mirror offset increased by 0.15 μm, (right) mirror offset increased by 0.30 μm

3 ESPI in the Study of Musical Instruments in Our Labs

Several musical instruments have been studied in our labs utilizing the time averaging ESPI method or other similar optical techniques. Among the string instruments that have been studied are the Ancient Greek Lyra Chelys [2, 24], Cretan Lyra [7, 10, 23], Pontic Lyra, Bouzouki [8], Baglamas, Tzuras, Violin, Mandolin and classical Guitar. Furthermore, percussion instruments have been studied (Bendir, Snare Drum, Bell) and Wind instruments (Zournas). Four representative studies for a Cretan Lyra, a Bendir, a classical Guitar and the ancient Greek lyra Chelys that we have performed are selected and presented in the following subsections, for the demonstration of the developed vibration analysis and measurement techniques.

3.1 The Cretan Lyra

The Cretan lyra is a Greek pear-shaped, three-stringed bowed musical instrument, of central importance for the traditional music of Crete and other Greek Aegean islands. The Cretan lyra is considered to be the most popular surviving form of the medieval Byzantine lyra, an ancestor of most European bowed instruments. In Fig. 6 the characteristics of a Cretan lyra geometry are depicted [7].

The lyra is held vertically on the musician's lap, in the same way as a small viol, rather than being placed under the chin of the player like a violin. For normal right-handed playing, the player's right hand holds the bow. The strings are stopped by pressing the fingernails of the player's left hand against the side of the string.

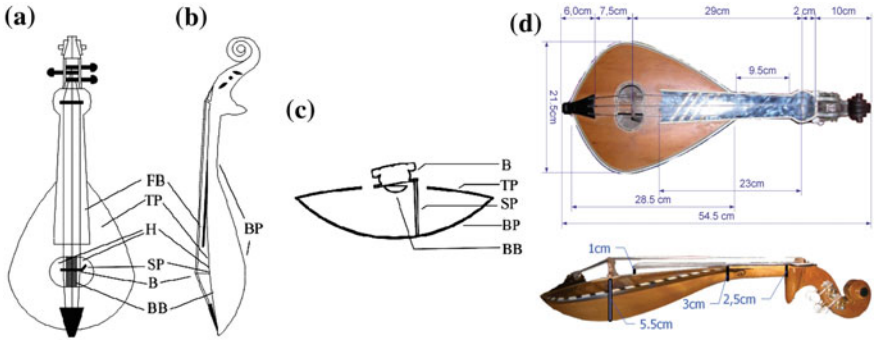


Fig. 6 The Cretan lyra. **a** Front view, **b** left view and **c** cross section view at the bridge: *TP* marks top plate; *BP* back plate; *B* bridge; *SP* soundpost; *FB* fingerboard; *H* holes, **d** typical dimensions

3.1.1 ESPI in the Study of Cretan Lyra

The acoustical properties of bowed musical instruments and especially those belonging in the violin family, have been extensively studied during the past decades. Several different techniques have been employed: holographic interferometry [1, 9, 13, 16, 19], spectral response techniques [1, 9, 13, 19] and spectral emission techniques. From these studies it became evident that the individual parts of the instruments play an important role in the emitted sound [1, 9].

A Cretan lyra acoustical study was performed for the first time [7], since previous studies were limited in individual parts of the instrument, like for example, the top plates [3, 11]. The vibrational characteristics and behavior of musical instruments and their individual parts is crucial to the emitted sound and instrument's performance.

In Fig. 7, time-average ESPI experimental vibration results of the main normal modes of two Cretan lyras of different period, are presented. The first one is a pear-shaped Cretan lyra of 17th century and the other a contemporary one.

The differences in the normal-frequencies observed are attributed mostly to the different geometrical characteristics and shapes. The pear-shaped lyra of the 17th century is in general smaller in all characteristic dimensions and as a result its resonance frequencies are higher for the same characteristic normal mode in comparison with the contemporary one.

3.2 Bendir

The percussion instruments are the oldest musical instruments and their acoustic properties are extensively studied worldwide but it is a fact that only a few and non-systematic studies have been performed for the traditional percussion instruments in Southeastern Mediterranean region. These percussion instruments mainly consist of

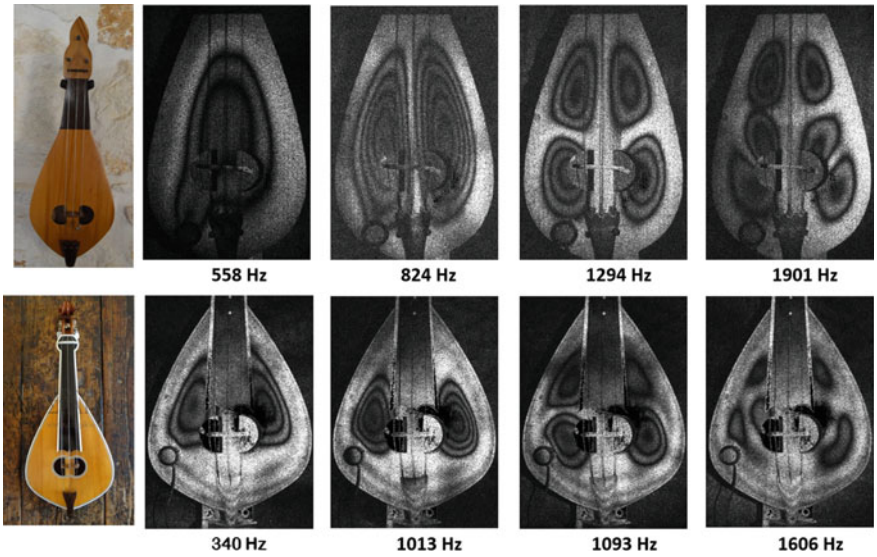


Fig. 7 Top row: pear-shaped lyra 17th century A.D., bottom row: contemporary Cretan lyra

a cylindrical frame on which a membrane, of an animal skin or of a plastic film, is stretched. The normal ways of oscillating this membrane play an important role to the characteristics of the finally broadcasted sound from the instrument. The main percussion instrument of this type is the bendir.

The bendir is a frame drum with a wooden frame (diameter ~ 35–45 cm) and a membrane (typically skin leather). It has no jingles, but most often has a snare (usually made of gut) stretched across its head, which when the drum is struck with the fingers or palm gives the tone a buzzing quality. It creates different tones according to the spreading of the shock waves moving across the skin itself. The drum is played, kept vertical, by inserting the thumb of the left hand in a special hole, shown in Fig. 8, in the frame.

The characteristic eigenfrequencies and the corresponding modes of the membrane of Bendir were studied. The structural dimensions of Bendir analyzed in this study vary, but the most common values of the diameter of this cylindrical instrument are in the range of 35–45 cm. It is commonly manufactured by the assembling of a cylindrical wooden frame to the natural leather or synthetic membrane and during play it is held in a vertical position and is supported by the thumb of the left hand in special holes on its frame. The instrument was experimentally studied using a 532 nm Nd:YAG laser source in a specially developed ESPI setup. Important mechanical characteristics of the instrument, like the winding tensions and the membranes material properties, that affect its acoustic behavior, are modified and their influence on the modes is monitored and recorded. The experimental results are compared to numerical results based on a Finite Element Analysis (FEA) performed for the instrument Computer Aided Design (CAD) model [4, 15, 21, 22].

Fig. 8 The bendir instrument

3.2.1 Vibrational Analysis of Bendir

The comparison of the experimental to the simulation results is presented in Fig. 9. The model is prestressed by the help of a pressure load applied at the inner surface of the supporting cylindrical frame of the membrane in order to simulate the low tuning.

After the first set of ESPI measurements, the instrument is further medium and then highly tuned. The uniform pressure load is further raised to simulate the intermediate and finally the high tuning. In Figs. 10 and 11 are presented the results for medium and high tuning respectively, where a satisfactory agreement between experiment and simulation is observed regarding the eigenmodes and the corresponding eigenfrequencies.

3.2.2 Effects of Bendir Membrane Temperature Changes

The temperature of the Bendir's membrane, made of calf skin, was controllable raised in relation to the environmental temperature and the change in the resonance frequency for four different normal modes was recorded using the ESPI. The experimental Set-up used is presented at the right of Fig. 12 and consists of two Halogen Lamps (of 150 W each) for the temperature increase and a type K Thermocouple connected to a multimeter for the temperature measurement. The influence of the temperature changes to the resonance frequency for the four normal modes selected is presented at the graph of Fig. 12.

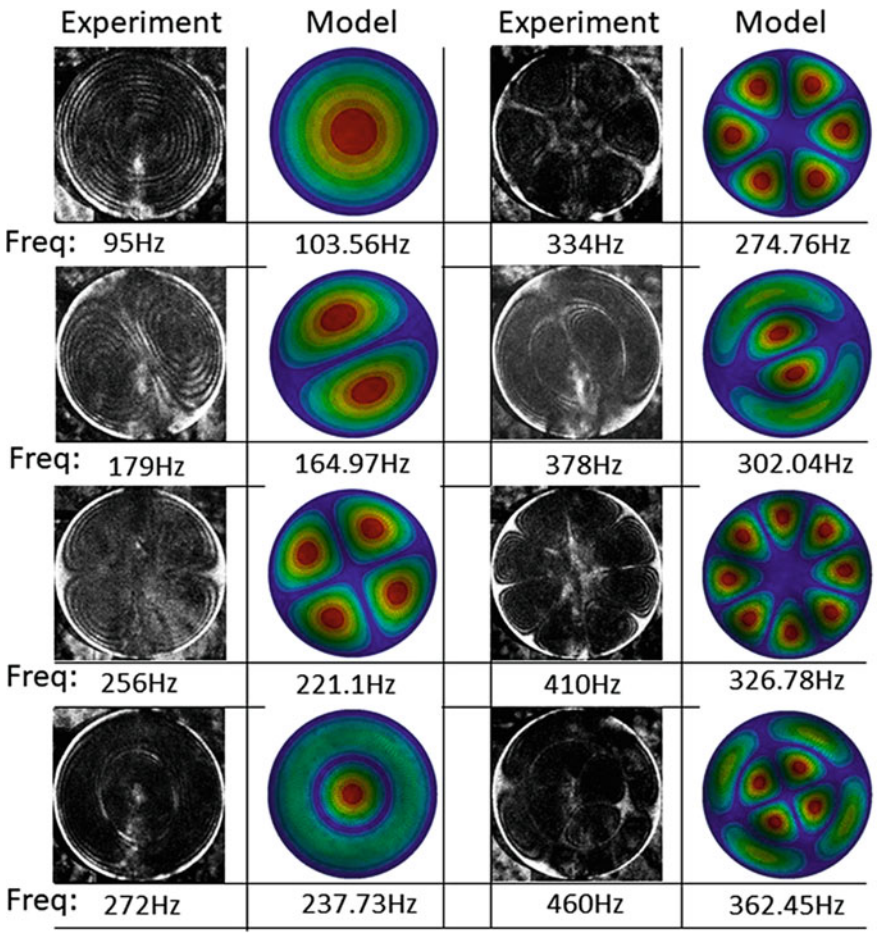


Fig. 9 Experimental results using ESPI and simulated results for the bendir in low tension pressure (low tuning)

3.3 Using ESPI for the Tuning of Guitar Top Plates

Tuning is achieved by removing a few grams of mass from the bracing. The same sound excitation parameters are applied for all top plates presented in Fig. 13. The depicted guitar top plate in Fig. 13 was manufactured by Koukourigou Bros. Even an experienced organ maker using tap tones cannot tune a new top plate very close to another that is characterized as prototype. ESPI is the best method to do that with extremely high accuracy.

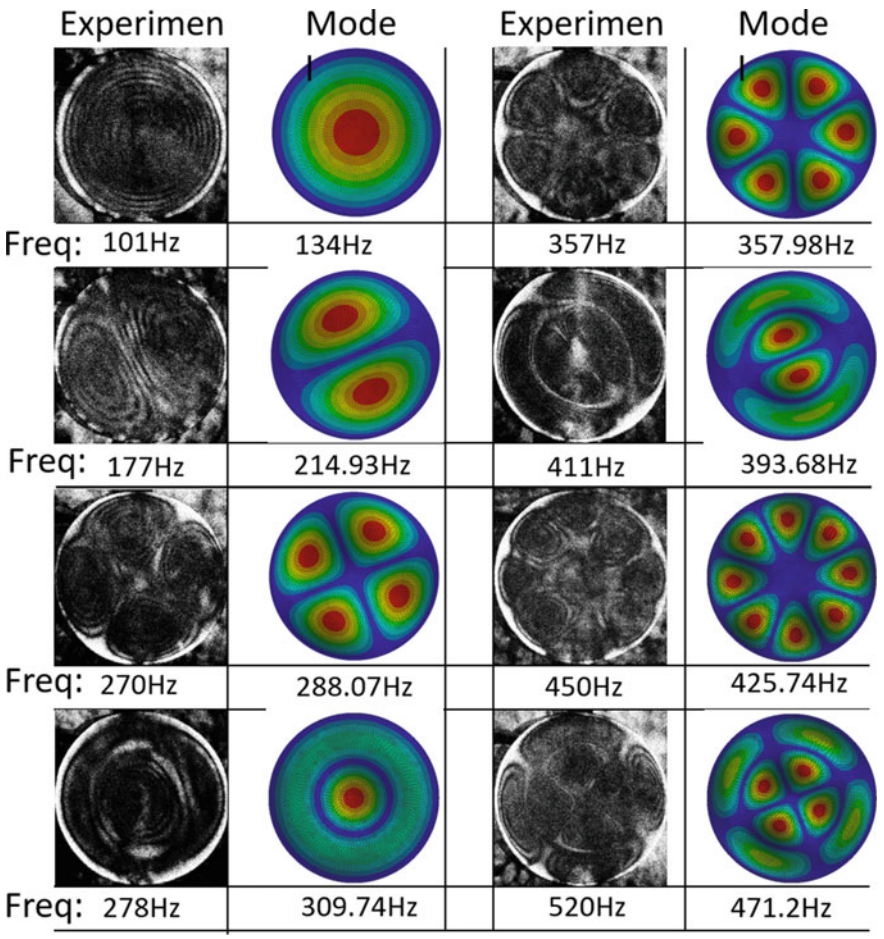


Fig. 10 Experimental results using ESPI and simulated results for the bendir in medium tension pressure (medium tuning)

3.4 The Ancient Greek Lyra Chelys—A Tortoise-Shell Lyra

The ancient Greek lyra Chelys is depicted on hydria (water pots) dated from the late 8th century B.C. Its origin in mythology is attributed to the god Hermes, who was the first one to craft such an instrument according to the 4th Homeric Hymn to Hermes. A faithful reconstruction, presented in Fig. 14, was performed using materials and tools available in Greek antiquity: sound box made using the carapace of the Greek tortoise Testudo Marginata (back) and ox leather (front), wooden parts made of oak wood, strings derived from dry sheep intestine.

The instrument was tuned according to the Phrygian mode which consists of the musical notes E3 (164 Hz), F3 (174 Hz), G3 (195 Hz), A3 (220 Hz), B3 (246 Hz),

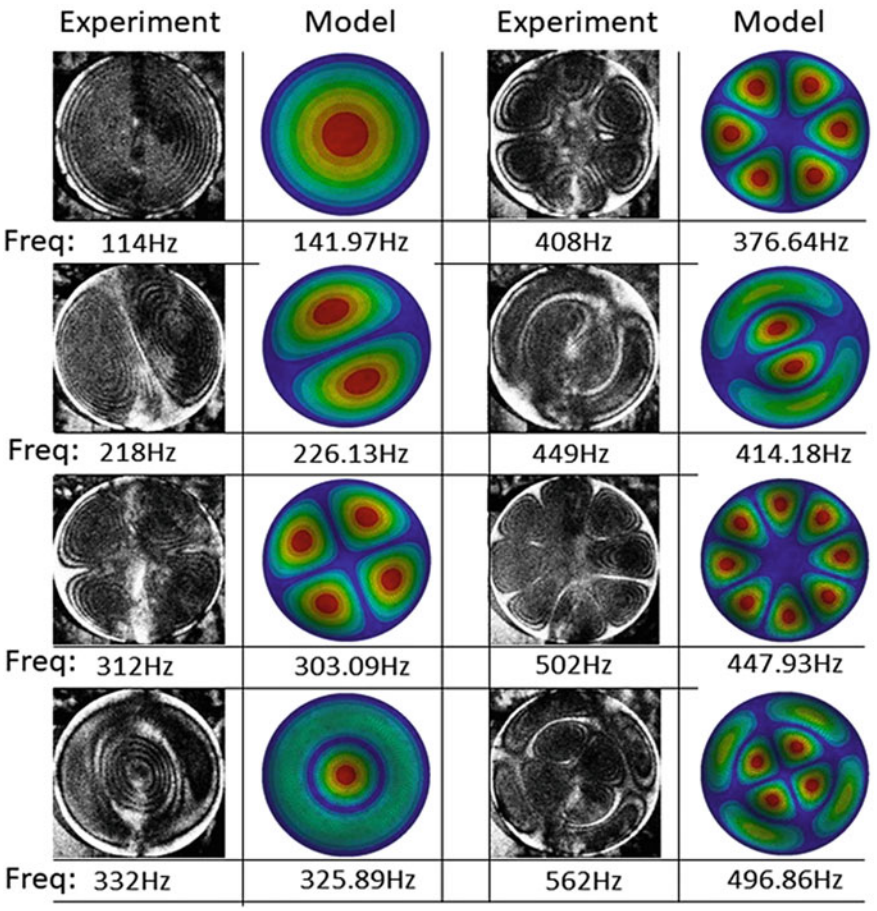


Fig. 11 Experimental results using ESPI and simulated results for the bendir in high tension pressure (high tuning)

C4 (261 Hz) and D4 (293 Hz). Recordings were made in a studio (RT_{60}: 0.37 s in the 500 Hz octave band), with a flat frequency response condenser studio microphone placed 20 cm away from Chelys, outside the near sound field. Individual notes were played in a controlled manner by an experienced player; strings stuck by a plectrum. The characteristic recorded sound is shown in Fig. 15.

Characteristic EPSI vibration results are presented in Fig. 16 [24].

With regard to the impulse response technique, an impulsive excitation was made by an impact hammer with feedback for the excitation force. Accelerometer was used to measure the object's response. Excitation and response signals were used to extract the dimensionless cross transfer frequency function, H2:

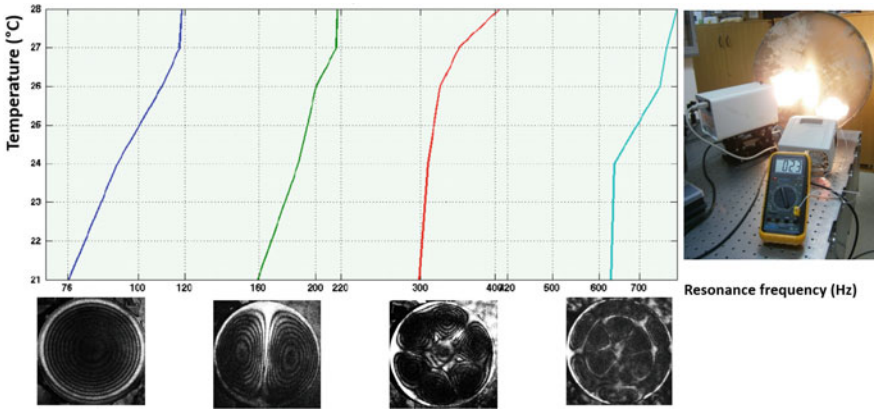


Fig. 12 Effects of temperature changes on the Bendir membrane

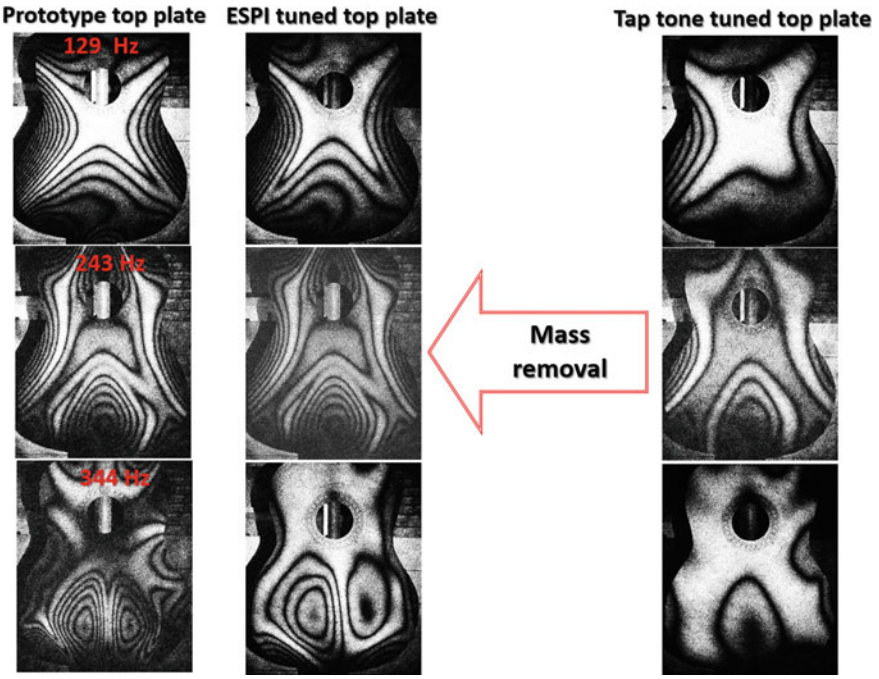


Fig. 13 The sound excitation parameters applied at top row: 129 Hz, middle row: 243 Hz, bottom row: 344 Hz

$$H2 = \frac{\left|\int_{-\infty}^{+\infty} y(t)e^{-i2\pi ft}dt\right|^2}{(\int_{-\infty}^{+\infty} x(t)e^{-i2\pi ft}dt)(\int_{-\infty}^{+\infty} y(t)e^{-i2\pi ft}dt)^*} \tag{10}$$

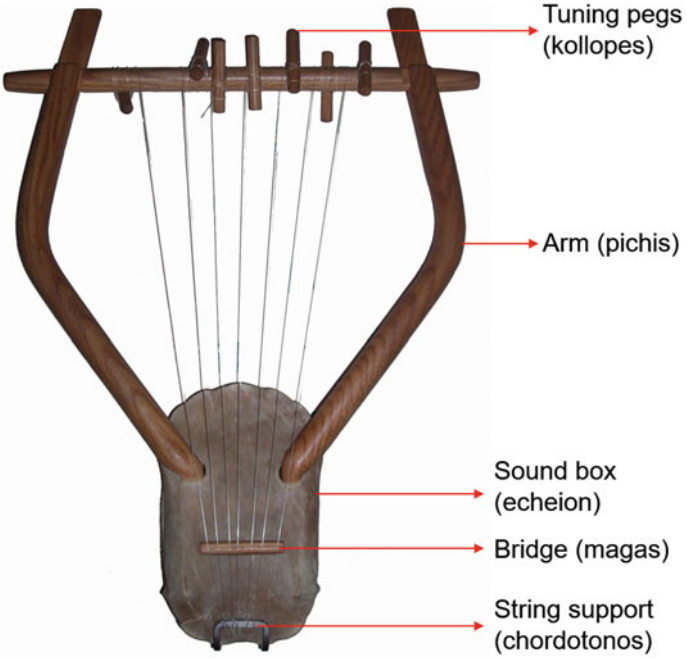


Fig. 14 Reconstruction of the ancient Greek lyra Chelys

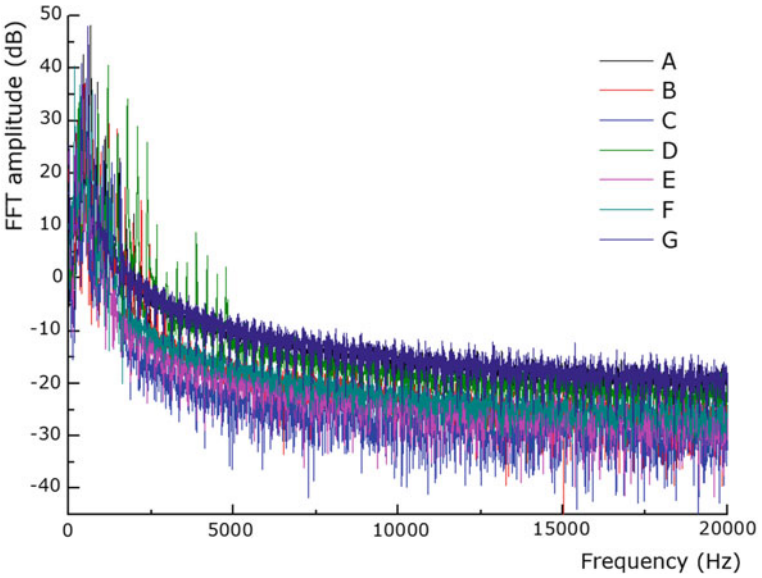


Fig. 15 The recorded sound from the reconstructed ancient lyra Chelys up to 20 kHz

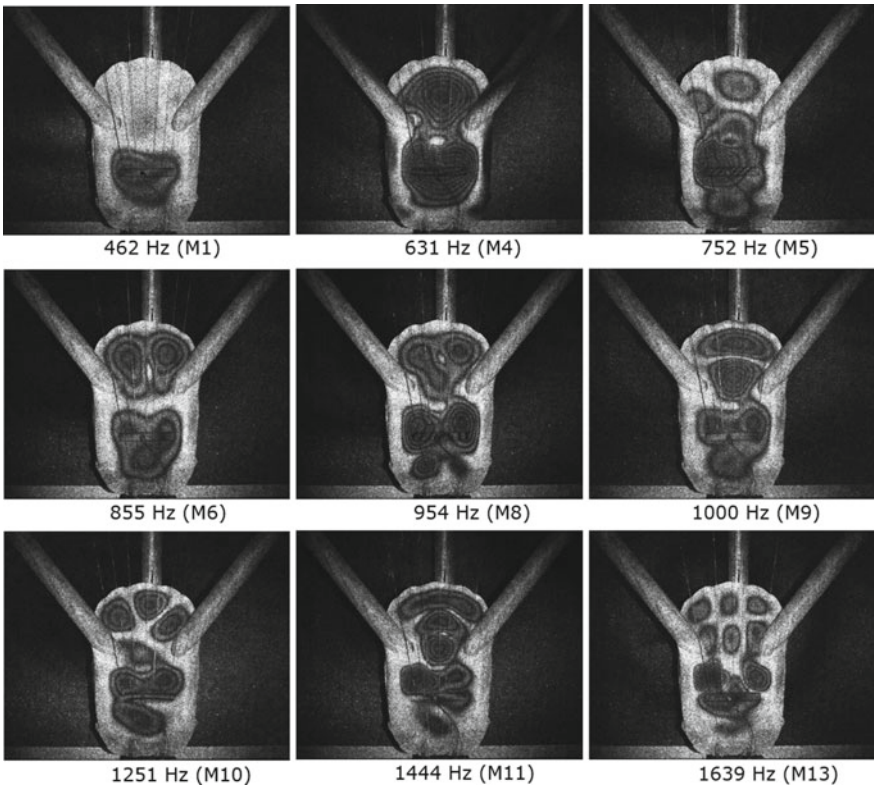


Fig. 16 Selected ESPI results from the reconstructed ancient lyra Chelys

where:

$y(t)$ accelerometer detection signal
$x(t)$ impact hammer excitation acceleration signal.

The impulse response frequency cross transfer function H2, in relation to the frequency according to the ESPI results, is presented in Fig. 17.

The radiated sound is mainly concentrated in frequencies up to 2 kHz, while the highest amplitudes are found in harmonics ranging from 400 to 800 Hz approximately. This concentration of spectrum energy, lies within the typical frequency range of mesophone human voice. This strengthens historical evidence, which suggests that the Chelys was most likely used as an accompaniment to the human voice.

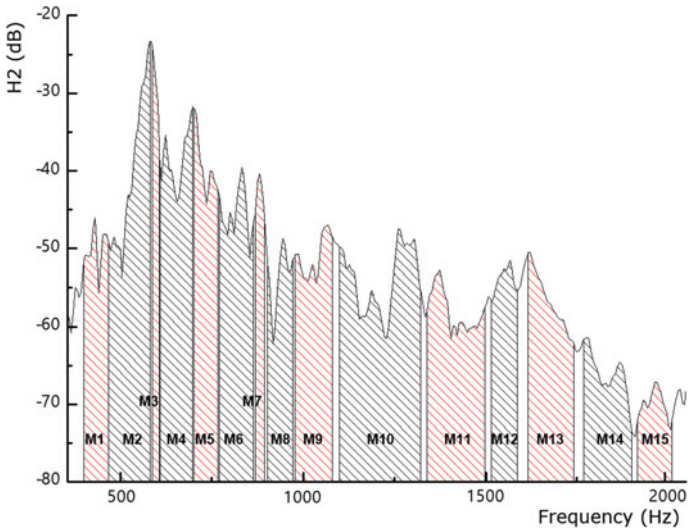


Fig. 17 The reconstructed Chelys impulse response frequency cross transfer function (solid line) divided into frequency regions. Each vibration mode appears according to ESPI results

4 Conclusions

The capabilities of laser-based optical techniques to investigate the vibrational characteristic of musical instruments is presented in this study. More emphasis is given in the use of the ESPI technique. The mathematical description of the time-average ESPI, the reading of the contour lines and the overcoming of the limitations of amplitude and phase vibration are analyzed. Furthermore, four representative studies using the ESPI for a Cretan Lyra, a Bendir, a classical Guitar and the ancient Greek lyra Chelys, are presented and demonstrate the capabilities of the method to determine normal modes and vibration amplitude distribution in whole field with ultra-high resolution, resonance frequencies and the phase of vibration.

References

1. Backus J (1977) The acoustical foundations of music. W.W. Norton & Co, New York
2. Bakarezos E, Vathis V, Brezas S, Orphanos Y, Papadogiannis NA (2012) Acoustics of the Chelys—an ancient Greek tortoise-shell lyre. Appl Acoust 73:478–483
3. Bakarezos M, Gimnopoulos S, Brezas S, Orfanos ϒ, Maravelakis E, Papadopoulos CI, Tatarakis M, Antoniadis A, Papadogiannis NA (2006) Vibration analysis of the top plates of traditional Greek string musical instruments. In: Eberhardsteiner J, Mang HA, Waubke H (eds) The thirteenth international congress on sound and vibration, Vienna, Austria
4. Bakarezos E, Orphanos Y, Kaselouris E, Dimitriou V, Tatarakis M, Papadogiannis NA (2018) Application of laser interferometric methods for studying traditional and ancient Greek music

instruments. In: Proceedings of the 9th panhellenic conference acoustics 2018, Patras, Greece
5. Baker LR (1983) Holographic and speckle interferometry. Optica Acta Int J Opt 30:1041
6. Castellini P, Revel GM, Tomasini EP (2006) Laser Doppler vibrometry: development of advanced solutions answering to technology's needs. Mech Syst Signal Pr 20:1265–1285
7. Chartofylakas L, Bakarezos E, Orphanos Y, Papadogiannis NA (2008) Connecting the acoustic characteristics of the structure of the Cretan lyra with the quality of the emitted sound. In: Proceedings of the 4th panhellenic conference acoustics 2008, Xanthi, Greece, pp 182–191
8. Chartofylakas L, Floros A, Bakarezos E, Papadogiannis NA (2010) Acoustic and sound analysis of the instruments of the bouzouki family. In: Proceedings of the 5th panhellenic conference acoustics 2010, Athens, Greece, pp 422–428
9. Fletcher NH, Rossing TD (1999) The physics of musical instruments. Springer, New York
10. Gimnopoulos S, Bakarezos M, Vathis V, Chartofylakas L, Brezas S, Orphanos Y, Maravelakis E, Papadopoulos CI, Tatarakis M, Antoniadis A, Papadogiannis NA (2006) Acoustic and interferometric analysis of the Cretan lyre. In: Proceedings of the 3rd panhellenic conference acoustics 2006, Heraklion, Greece, pp 239–246
11. Gimnopoulos S, Kouzoupis S, Bakarezos M, Orfanos Y, Tatarakis M, Papadogiannis NA (2004) Vibrational analysis of Greek string instruments top plates: preliminary experimental results using mechanical and laser optical techniques. In: Proceedings of the 2nd panhellenic conference acoustics 2004, Thessaloniki, Greece, pp 93–100
12. Huang C-H, Ma C-C (2001) Experimental and numerical investigations of resonant vibration characteristics for piezoceramic plates. J Acoust Soc Am 109:2780–2788
13. Jansson E, Molin N-E, Saldner HO (1994) On eigenmodes of the violin—electronic holography and admittance measurements. J Acoust Soc Am 95:1100–1105
14. Jones R, Wykes C (1989) Holographic and speckle interferometry. Cambridge University Press, Cambridge
15. Kokkinakis E (2013) Finite element modeling simulation of the acoustic behaviour and characteristics of a drum. BSc Thesis, Department of Music Technology & Acoustics Engineering, TEI Crete
16. Lökberg O (1984) ESPI—the ultimate holographic tool for vibration analysis? J Acoust Soc Am 75:1783–1791
17. Molin N-E (1999) Applications of whole field interferometry in mechanics and acoustics. Opt Lasers Eng 31:93–111
18. Rastogi PK (ed) (2001) Digital speckle pattern interferometry and related techniques. Wiley, Chichester
19. Runnemalm A, Molin N-E, Jansson E (2000) On operating deflection shapes of the violin including in-plane motions. J Acoust Soc Am 107:3452–3459
20. Sharp B (1989) Electronic speckle pattern interferometry (ESPI). Opt Lasers Eng 11:241–255
21. Sidiras G, Kokkinakis E, Orphanos Y, Bakarezos E, Kaselouris E, Dimitriou V, Papadogiannis NA (2014) Vibrational features of traditional percussion music instruments using laser and numerical simulations. In: Proceedings of the 7th panhellenic conference acoustics 2014, Thessaloniki, Greece, pp 12–19
22. Sidiras G (2013) Investigation of vibrational features in traditional percussion music instruments by using optical interferometry techniques, BSc Thesis, Department of Music Technology & Acoustics Engineering, TEI Crete
23. Theodosopoulou I, Chartofylakas L, Bakarezos M, Orphanos Y, Papadogiannis NA (2009) The Cretan lyre: an ethnomusicological and music acoustics approach. In: Proceedings of the CIM09, pp 172–174
24. Vathis V, Bakarezos E, Orphanos Y, Papadogiannis NA (2008) Acoustic study of the faithful reconstruction of the ancient Greek Chelys lyre. In: Proceedings of the 4th panhellenic conference acoustics 2008, Xanthi, Greece, pp 173–181
25. Wang CP (1988) Laser Doppler velocimetry. J Quant Spectrosc Radiat Transfer 40:309–319
26. Wang W-C, Hwang C-H, Lin S-Y (1996) Vibration measurement by the time-averaged electronic speckle pattern interferometry methods. Appl Opt 35:4502–4509

Shock Wave Characteristics in the Initial Transient of an Organ Pipe

Jost Leonhardt Fischer

No enjoyment is transitory; it leaves an impression behind; that impression is abiding, and whatever is done earnestly and industriously imparts even to the mere looker-on a hidden power whose effects spread farther than we can ever know.
Johann Wolfgang von Goethe, Wilhelm Meisters Apprenticeship, 1795/96, 5. Book, Chapt. X

Abstract A new approach to investigate the role of pressure waves in the initial transient of an organ pipe is presented. By numerical simulations solving the compressible Navier-Stokes equations with suitable boundary and initial conditions, it is possible to retrace the generation, propagation, reflection, damping and radiation of sound waves. The focus is on the contribution of occurring pressure wave fronts in the initial transient that show shock wave characteristics, in particular their role at the formation process of the sound field inside the organ pipe's resonator. Utilizing spectral analysis as well as extended visualization methods, a wide range of aspects of the dynamics of the initial transient of an organ pipe is discovered. In particular the damping processes in the resonator which are nonlinear are analyzed and discussed in detail. The numerical approach presented in this case study, allows to study the initial transient of an organ pipe with an extraordinary level of precision, thereby helping to understand the underlying first principles of the sound field formation and the mutual interaction of the jet's flow field and the sound field inside organ pipes and similar wind instruments that produce complex sounds listeners find both interesting and joyful. Animations of the temporal and spatial development of relevant physical quantities like pressure, turbulent kinetic energy, vorticity and velocity magnitude calculated in the numerical simulations are provided as supplementary material.

Electronic supplementary material The online version of this chapter (https://doi.org/10.1007/978-3-030-02695-0_13) contains supplementary material, which is available to authorized users.

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

R. Bader (ed.), *Computational Phonogram Archiving*, Current Research in Systematic Musicology 5, https://doi.org/10.1007/978-3-030-02695-0_13

1 Introduction

In this chapter, numerical investigations and measurements of the processes in the initial transient of an organ pipe are discussed. The question that emerges immediately is what this topic can contribute to ethnomusicology. The case study presented here helps to discover the complex dynamics in organ pipes as well as in similar end-blown wind instruments, for instance the *Turkish Ney*, the *Persian Ney*, the shakuhachi and the recorder. It may helps to understand the different kinds of playing techniques of the addressed instruments, especially the blowing techniques at the beginning. This in turn is an important detail of the cultural aspects of making music which are explored in ethnomusicology. The initial transient is one of the persistent secrets in musical acoustics. In particular the dynamics in the initial transient, the interactions of the instrument's jet flow field and the sound field has not yet been fully understood. In the following, I shall first point to some previous works on the topic of transients in aerophones.

Fletcher gave a set of coupled nonlinear differential equations for the normal modes interacting through the air jet driving the pipe [6]. Verge et al. did flow visualizations of the initial transient in a small recorder-like flue organ pipe and showed that the various stages of the jet formation are related to measurements of the acoustic response of the pipe [14]. Yoshikawa investigated attack transients in organ pipes by slow-motion pictures from the smoked-jet visualization with a high-speed digital video camera [15]. The author's addressed did not observe or refer to the occurrence of shock waves or pressure waves that have shock wave characteristics in the initial transient of the wind instruments under investigation.

However, Campbell [2] noted that nonlinear effects in the propagation of high amplitude sound waves can lead to the development of shock waves in trumpets and trombones, with important musical consequences. Hirschberg observed the formation of shock waves at fortissimo level in trombones [7]. The focus of the work presented here is on pressure wave fronts in the initial transient that show shock wave characteristics and their role in the processes of sound generation. Of particular interest will be their contribution to the formation of the sound field in the resonator of the instrument.

The investigations were carried out using numerical simulations of a stopped wooden organ pipe. The case study discovers the general effects of emerging pressure wave fronts in the resonator of an organ pipe during the initial transient process which show shock wave characteristics and their role at the formation process of the sound field in particular. The motivation is to understand the first principles of the complex dynamics of sound generation in an organ pipe and similar wind instruments. It is shown that the observed pressure wave fronts, in the following named shock waves, or shocks, play a key role in the initial transient process.

The original wooden organ pipe used in the study as a template for the numerical model was provided by the German organ builder Alexander Schuke Orgelbau Potsdam GmbH [13], cf. Fig. 1a.

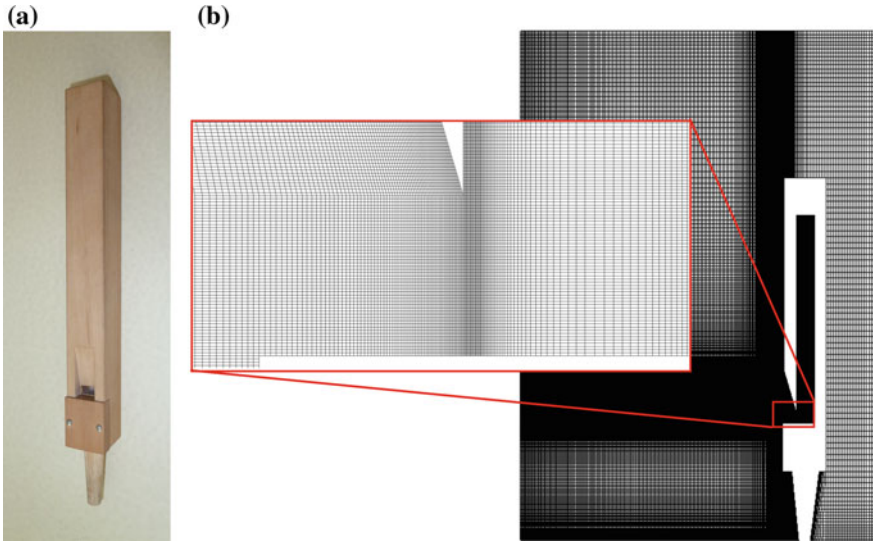


Fig. 1 **a** Stopped wooden organ pipe with a quadratic cross-section, built and provided for measurement use by organ builders Schuke Orgelbau Potsdam GmbH. **b** Implementation of the organ pipe and the surrounding space into a 2D computational grid. The detail gives an impression of the mesh size in the cut-up region

The work has the following structure. The first part gives some general notes about the procedure of successful implementation and run of advanced numerical simulations such as presented here. Specifications of the case that is studied are outlined.

The constitutive equations are given, relevant fluid mechanical characteristic numbers are discussed and their estimations regarding the case are addressed. Furthermore, estimations of the Kolmogorov-scales for the case under investigation are discussed as well as the consequences concerning the decisions of the grid size of the calculation grid to be used. Software and hardware decisions are stated. The configuration of thermo-physical properties needed in the study are addressed as well as the values for boundary and initial conditions of the physical quantities to be calculated. In addition the turbulence model to be favored is introduced and the mesh which is utilized is presented.

In the second part, methods of analysis and the results of the investigations are subjects of discussion. The analysis starts with a spectral analysis of the time series sampled in the resonator. The occurrence of pressure wave fronts with shock wave characteristics in the initial transient and their propagation along the longitudinal axis in the resonator is depicted and discussed. The observed nonlinear dissipative behavior of the named wave fronts manifest in both velocity and amplitude data is examined. With enhanced methods and techniques of visualization, new insights into the dynamics of the initial transient are gained, in particular the interaction between initial pressure wave fronts and sound field formation inside the organ pipe.

2 General Notes on Numerical Implementation and Numerical Simulation

The interaction between flow field and sound field, the sound generation and the sound propagation are described by the compressible Navier-Stokes equations [9, 12]. For a successful numerical implementation of the given set-up, which is a wooden stopped organ pipe whose internal air volume is set to vibration by a blowing mechanism, one has to apply the compressible Navier-Stokes equations with appropriate initial and boundary conditions. The set of equations have to be solved on an suitable computational grid, the numerical space which is called mesh.

The numerical treatment of compressible problems announced here is still an advanced task. The main difficulties arise from reproducing the interactions between the flow field and the sound field [4] because of the different scales the flow velocity and the particle velocity act on. Numerical simulations allow to study the dynamics of inherent fluid mechanical structures like vortices, jets as well as the generation of sound waves, their propagation in the resonator and their radiation into the free space simultaneously. In principle, a successful modeling and simulation process can be divided into the following sections: Physical previews, Pre-processing, Processing and Post-Processing. The sections include the following sub-tasks and relate to questions that need to be being answered appropriately:
Physical Previews:

- Which set of equations is constitutive for the given problem?
- Which are the characteristic fluid dynamical numbers to be taken into account?
- What are the scales of the problem?
- Software-decision.
- Hardware-decision.

Pre-Processing:

- How to write an appropriate mesh for the given case?
- Determine the relevant thermo-physical properties.
- Implement suitable initial and boundary conditions for each physical quantity to be calculated, e.g. pressure p, the velocity vector $\vec{U}$, temperature T, density ρ, turbulent kinetic Energy k, etc.
- Discretization schemes for the differential operators in the constitutive equations (del operator, Laplacian, time derivative, etc.) inclusive proper correctors.
- Select an appropriate turbulence model to model the energy transfer into and out of the sub-grid scales.
- Solver for the compressible fluid dynamical problem, determination of numerical schemes and their tolerances.
- Adequate matrix solvers.
- Configure relevant numerical parameters, e.g. numerical time step size, simulation time, write precision etc.
- Define suitable sample sets and probe points in the mesh for analysis.
- Parallelize the calculation.

- Take care of numerical stability parameters, e.g. Courant number.
- Control during simulation run time.
- Calculate additional physical quantities, e.g. vorticity, etc.

Post-Processing:

- Visualize the simulation.
- Analysis.

For more detailed information the reader is referred to the author's Ph.D. thesis [5]. The numerical simulations presented here were realized by using parts of the C++ toolbox OpenFoam-3.0.0 [10]. The libraries include customized numerical solvers as well as pre- and post-processing utilities for the solution of problems in continuum mechanics, 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. General aspects about pre-processing, run and post-processing are documented in the OpenFOAM User Guide as well as in the OpenFOAM Programmer Guide [10]. Some of the issues addressed which are of special importance are pointed out in the following.

2.1 Constitutive Equations

The set of constitutive equations one has to deal with contains: the continuity equation, Eq. (1), the momentum balance equations (compressible Navier-Stokes equations), Eq. (2), the energy balance equations, Eq. (3) with Fourier's law, Eq. (4), the equation for the total energy (internal and kinetic energy), Eq. (5), the calorimetric state equation (enthalpy), Eq. (6), the equation to closure the problem, here the relation between pressure and internal energy, Eq. (7) with the isentropic coefficient, Eq. (8):

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \boldsymbol{u}) = 0 \tag{1}$$

$$\rho \left(\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u} \cdot \nabla) \boldsymbol{u} \right) + \boldsymbol{u} \frac{\partial \rho}{\partial t} = -\nabla p + (\lambda + \eta) \nabla (\nabla \cdot \boldsymbol{u}) + \eta \boldsymbol{u} + \rho \cdot \boldsymbol{g} \tag{2}$$

$$\frac{\partial (\rho E)}{\partial t} + \nabla \cdot (\rho E \boldsymbol{u}) = -\nabla (p \boldsymbol{u}) + \nabla (\boldsymbol{\tau} \boldsymbol{u}) - \nabla \boldsymbol{q} \tag{3}$$

$$\text{with} \qquad \boldsymbol{q} = -\kappa \nabla T \tag{4}$$

$$E = e + \frac{1}{2} |\boldsymbol{u}|^2 \tag{5}$$

$$H = e + \frac{p}{\rho} = C_p T \tag{6}$$

$$p = (\gamma - 1)\rho e \tag{7}$$

$$\gamma = \frac{C_p}{C_V} \approx 1.4 \tag{8}$$

The given set of equations one has to discretize utilizing appropriate discretization schemes for each differential operator, e.g. backward for the time derivative operator or Gauss linear for the del operator, the divergence operator and the Laplacian, cf. [10].

2.2 Fluid Mechanical Characteristic Numbers

An important and necessary step is to estimate the relevant characteristic fluid mechanical numbers for the given set-up to fit the numerical implementation of the problem. The probably most important fluid mechanical characteristic number one has to estimate in order to solve the problem under review is the Reynolds number. The Reynolds number is the dimensionless ratio of inertial forces to viscous forces [12]. The forces are represented by the characteristic length L of the dynamic in the case and the absolute value of the characteristic velocity U of the flow field, on the one hand, and the kinematic viscosity on the other, cf. Eq. (9). The Reynolds number is used to valuate the transition of laminar to turbulent in a particular flow field, e.g. in a pipe flow with diameter d turbulent flow occurs if $Re_d > 4000$ [12]. The characteristic length of the case discussed here is given by the length of the cut-up of the organ pipe, $L = 5.5 \times 10^{-3}$ m which determines the jet flow area. The characteristic velocity is assumed as $U = |u_y| = 18$ m/s, the absolute value of the initial velocity profile of the jet at the windway. The kinematic viscosity $\nu = \mu/\rho$ of the medium air is about $\nu = 1.53 \times 10^{-5}\,\mathrm{m}^2/\mathrm{s}$. With these values the Reynolds number can be estimated as follows

$$Re = \frac{L \cdot U}{\nu} \approx 6470 \tag{9}$$

The Reynolds number obtained for the given problem indicates that the flow field in the mouth region of the organ pipe is of weak turbulence [12]. Note, that the Reynolds number estimates the flow characteristics relative to the viscous characteristics of the medium and not the characteristics of the sound field, which acts on a much faster time scale. Therefore the flow velocity has to take into account and not the speed of sound.

As a further important characteristic number the Strouhal number is discussed. The Strouhal-number is the dimensionless ratio between an aeroacoustical quantity, the so-called vortex shedding frequency f, and the fluid mechanical quantities, the characteristic length L and the characteristic velocity U of the flow [12]. The vortex shedding frequency in the mouth region is caused by the oscillations of the jet of

Table 1 Implemented thermo-physical properties

Property	Value	Unit
Molecules	1	
Molar mass M	28.9	g/mol
Specific heat capacity (p=*const*) C_p	1007	J/(kg K)
Dynamic viscosity μ	1.8×10^{-5}	Pa s

the working organ pipe. Here the hypothesis is that the vortex shedding frequency corresponds to the fundamental frequency of the real operating organ pipe $f \approx 750\,\text{Hz}$, which one obtains by measurements. The Strouhal number is estimated by

$$St = \frac{L \cdot f}{U} \approx 0.23 \tag{10}$$

The estimated Strouhal number corresponds to values one would expect for the problem considered.

The Prandtl number is defined as the ratio of kinematic viscosity, also known as momentum diffusivity, to the thermal diffusivity, cf. Eq. (11). The thermal diffusivity of the medium air at normal conditions ($p = 1013.25\,\text{hPa}$, $T = 293\,\text{K}$) is given by $\alpha = \kappa/(\rho \cdot C_p)$. Hereby $\kappa = 0.0257\,\text{W}/(\text{m} \cdot \text{K})$ is the thermal conductivity, cf. Eq. (4) and C_p the specific heat capacity at constant pressure conditions, cf. Eq. (6) and Table 1. The thermal diffusivity is calculated as $\alpha = 1.9 \times 10^{-5}\,\text{m/s}$. This leads to the estimation of the Prandtl number of

$$Pr = \frac{\nu}{\alpha} \approx 0.72 \tag{11}$$

Last but not least the Mach number is important to estimate. The Mach number is defined as the ratio of the characteristic velocity U to the speed of sound c_0. The Mach number in the case being considered is estimated to

$$Ma = \frac{U}{c_0} \approx 0.052 \tag{12}$$

with $c_0 = 343\,\text{m/s}$ the speed of sound at normal conditions. Note, that the occurrence of pressure waves that show shock wave characteristics in the initial transient is not yet expected taking this estimation.

2.3 Kolmogorov-Microscales, Grid Size

The second essential step is to estimate the Kolmogorov microscales [8] needed to evaluate the chosen computational grid sizes as well as the resolution of the numerical time step size. The Kolmogorov length scale η and time scale τ_η are the microscales

where the viscosity dominates and the turbulent kinetic energy is dissipated into the heat bath. The estimation of the Kolmogorov microscales is as useful as necessary to find the optimal grid size and time scale to solve the problem numerically. The Kolmogorov microscales are defined as

$$\eta = \left(\frac{\nu^3}{\epsilon}\right)^{1/4}, \qquad \tau_\eta = \left(\frac{\nu}{\epsilon}\right)^{1/2} \tag{13}$$

where $\epsilon = U^3/L$ is the average rate of dissipation of turbulence kinetic energy k per unit mass. The Kolomogorov microscales are estimated at $\eta = 7.62 \times 10^{-6}$ m and $\tau_\eta = 3.8 \times 10^{-6}$ s. The time increment of the numerical simulations is set to $\delta t = 10^{-8}$ s to ensure numerical stability, and therefore smaller by the factor 380 than the Kolomogorov time scale requires. Note that numerical stability of explicit time integration schemes, which are, inter alia, utilized in the numerical simulations, is given by the Courant-Friedrichs-Lewi condition (CFL-condition)

$$\mathrm{CFL} = \frac{U \cdot \delta t}{\delta x} \leq C_{max} = 1 \tag{14}$$

The smallest grid sizes of the created mesh are $\delta x = 1 \times 10^{-4}$ m and $\delta y = 2 \times 10^{-4}$ m. Hence $\mathrm{CFL}_{\max} = 0.35 \leq 1$ and therefore Eq. (14) is satisfied by choosing δt as mentioned.

Compared with the Kolmogorov length scale the smallest grid sizes are too large by the factor 14. That means, that turbulent structures smaller than the grid size cannot be resolved by the mesh. This fact makes a turbulence model necessary, which models the energy transfer into and out of the sub-grid scales, cf. Sect. 2.7.

2.4 Software and Hardware

The numerical simulations were realized by using parts of the C++ toolbox OpenFoam-2.1 [10]. 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.

The numerical simulations were calculated on the HPC-Cluster at the University of Hamburg using 16 nodes with a total of 256 CPUs. The OpenFOAM solver `rhoPimpleFoam` for compressible problems was used. An amount of data of about 81 GB per simulation run was generated. The run time of one simulation was 5.5 h. The simulation times were up to $t_s = 100$ ms.

Visualizations of the numerical simulations were generated using the open source multi-platform data analysis and visualization application ParaView-4.1 [10].

The data analysis and signal processing were programmed using MATLAB®.

2.5 Thermo-physical Properties

The thermo-physical properties one has to take care of are the molar mass of the medium which is the gas air, the specific heat capacity and the dynamic viscosity. An often used simplification is the assumption of air as a perfect, one-atomic gas. The values of the thermo-physical properties which are used in our case are summarized in Table 1.

2.6 Boundary and Initial Conditions

One has to define the boundary conditions at the peripheries of the mesh, the wall properties of the organ pipe and the inlet conditions at the organ pipes languid. The organ pipe's surfaces are considered as acoustically inert (boundary condition: no slip). The peripheries of the mesh are configured as open. This means that the radiated sound as well as all other physical quantities can propagate through the boundaries without any restrictions.

The initial conditions ($t_0 = 0\,\mathrm{s}$) for the physical properties to be calculated are implemented, e.g. the velocity profile of the flow field at the windway, which is assumed to be uniform (hat-profile): $\boldsymbol{u}(x, y, z, t_0) = 18\,\mathrm{m/s} \cdot \mathbf{e_y} = u_y$, the value of the initial pressure field $p(x, z, y, t_0) = 1013.25\,\mathrm{hPa}$, the value of the initial temperature field $T(x, y, z, t_0) = 293\,\mathrm{K}$. The boundary conditions at the organ pipe's inner walls are modeled by an appropriate wall function for the turbulent kinetic energy k. For the pressure at the walls the zero-gradient and for the velocity the no-slip boundary condition is chosen.

2.7 Turbulence Model

A spatial refinement of the mesh down to the Kolmogorov microscales would lead to an extensive increase of computing effort. To limit the costs and the effort a turbulence model is utilized. It models the transfer of turbulent kinetic energy k into and out of the sub-grid scales (SGS). The turbulent kinetic energy k can be split into a grid-scale term k_{GS} and a sub-grid scale term k_{SGS} as follows

$$k = \frac{1}{2}\overline{u_k u_k} = \underbrace{\frac{1}{2}\overline{u}_k\overline{u}_k}_{k_{GS}} + \underbrace{\frac{1}{2}(\overline{u_k u_k} - \overline{u}_k\overline{u}_k)}_{k_{SGS}} \tag{15}$$

using Einstein's summation convention for the spatial indices $k = 1, 2, 3$ of the velocity components u_k. As a suitable LES-Model for the SGS turbulent kinetic energy k_{SGS} (SGS-k model) a one-equation dynamic sub-grid scale model is selected [10]. The model equation for the transport of the turbulent kinetic energy k is given by Eq. (16). Further explanations of the mentioned terms are given by Eqs. (17)–(23). More detailed information about the turbulence model used can be found in the OpenFOAM User Guide as well as in the OpenFOAM Programmer Guide [10].

$$\frac{\partial(\rho k_{SGS})}{\partial t} + \frac{\partial(\rho\overline{u}_j k_{SGS})}{\partial x_j} - \frac{\partial}{\partial x_j}\left[\rho(\nu + \nu_{SGS})\frac{\partial k_{SGS}}{\partial x_j}\right] = -\rho\tau_{ij} : \overline{D}_{ij} - C_\epsilon\frac{\rho k_{SGS}^{3/2}}{\Delta} \tag{16}$$

with

$$k_{SGS} = \frac{1}{2}\tau_{kk} = \frac{1}{2}\left(\overline{u_k u_k} - \overline{u}_k\overline{u}_k\right) \tag{17}$$

$$-\rho\tau_{ij} : \overline{D}_{ij} = -\frac{2}{3}\rho k_{SGS}\frac{\partial\overline{v}_k}{\partial x_k} + \rho\nu_{SGS}\frac{\partial\overline{u}_i}{\partial x_j}\left(2\overline{D}_{ij} - \frac{1}{3}tr\left(2\overline{D}\right)\delta_{ij}\right) \tag{18}$$

$$\overline{D}_{ij} = \frac{1}{2}\left(\frac{\partial\overline{u}_i}{\partial x_j} + \frac{\partial\overline{u}_j}{\partial x_i}\right) \tag{19}$$

$$C_\epsilon = 1.05 \tag{20}$$

$$\Delta \quad \text{(Sauter mean diameter of the grid cell)} \tag{21}$$

$$\nu_{SGS} = C_k k_{SGS}^{1/2}\Delta \tag{22}$$

$$C_k = 0.07 \tag{23}$$

2.8 Mesh

The stopped wooden organ pipe, depicted in Fig. 1a, was produced and provided by the German organ builder Alexander Schuke Orgelbau GmbH [13]. The geometry of the organ pipe and its surrounding area is transferred into a structured 2D computational grid shown in Fig. 1b. The mesh size (length × width × depth) is (260 mm × 180 mm × 1 mm) with 254,342 mesh points, 505,170 faces and 126,000 hexahedra. The technique how to write an appropriate mesh file and how to generate a mesh can be found in the OpenFOAM User Guide [10].

3 Analysis and Results

3.1 Data Sampling

The numerical simulations produces an amount of data of ca. 500 GB which contains the field informations of the calculated physical quantities, pressure, velocity, density, temperature, turbulent kinetic energy, vorticity, etc, at each time step. For the analysis data sets at cross-section *cs_resonator* and *cs_jet* were sampled. The cross-section *cs_resonator* is displayed in Fig. 2 marked by the horizontal pink line, the cross-section *cs_jet* is marked by the vertical pink line. The data set *cs_resonator* has a spatial expansion of 100 mm and 1000 spatial sample points. The resolution is 0.1 mm. The cross-section *cs_jet* is 50 mm long and has 500 sample points. Overall 10,000 time steps with a resolution of $\delta t = 1 \times 10^{-5}$ s corresponding to the simulation time of 100 mm were sampled from the data of the numerical simulations. Note that the time resolution during the calculation of the numerical simulations was much finer with $\delta t = 1 \times 10^{-8}$. The sampled data were transferred into a time series by spatial integration over the lengths of the particular cross-section.

Figure 2 shows a snapshot of the numerical simulation, visualized at time step $t = 0.15$ ms. Color coded is the pressure field in the interval of 1011 hPa (light blue) to 1018 hPa (dark red). One observes pressure fronts reflecting at the inner walls of the organ pipes resonator, propagating to the closed end of the resonator. The origin of the pressure wave fronts is the jet flow which enters the stagnant air column in the resonator.

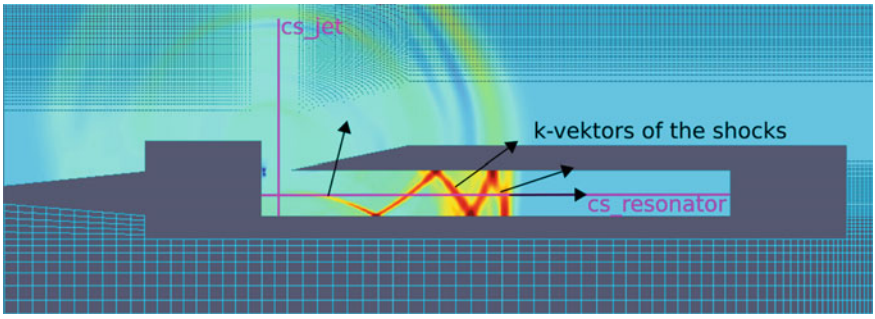


Fig. 2 Detail of the mesh of the modeled organ pipe. The mean resolution inside the resonator is $\delta x = 0.1$ mm, $\delta y = 0.1$ mm. The snapshot at time $t = 0.15$ ms shows initial pressure wave fronts reflecting at the inner walls of the resonator and propagating to the resonators upper end which is closed. The directions of k-vectors of the wave fronts are indicated by black arrows. Marked by the horizontal pink line along the longitudinal axis of the resonator is the position of the cross-section *cs_resonator*. Marked by the vertical pink line is the cross-section *cs_jet*, which is used for the analysis of the interaction of the jet's oscillations and the pressure field inside the resonator, discussed in Sect. 3.6

3.2 Spectral Analysis

The sound pressure level spectrum of the time series with the length of $t = 100$ ms has been calculated using Eq. 24. The reference pressure is the human auditory threshold $p_0 = 2 \times 10^{-5}$ Pa.

$$SPL = 20 \cdot log_{10} \left(\frac{p}{p_0} \right) \text{ dB} \tag{24}$$

The data have a sampling rate of $f_s = 100$ kHz caused by the temporal resolution of the data $\delta t = 1 \times 10^{-5}$ s written out from the numerical simulation. This corresponds to a resolution in frequency of $\delta f \approx \pm 6$ Hz of the spectrum. In Fig. 3 the SPL-spectrum of the spatial integrated time series is shown. One sees the fundamental frequency (first harmonic) at $f_1 \approx 757$ Hz followed by the next higher odd harmonics $f_3 = 2338$ Hz, $f_5 = 4090$ Hz, $f_7 = 5945$ Hz, $f_9 = 7911$ Hz, $f_11 = 9663$ Hz, $f_13 =$11,680 Hz. The deviation of the third harmonic to the theoretical value is 3%, which is an excellent result. The deviations of the fifth and seventh harmonic are 8% and 12% which are good results. In the SPL-spectrum one notices two significant peaks at very high frequencies of 2.144 and 2.533 kHz (Fig. 3). To give an explanation for this observation further methods have to apply.

3.3 Visualization

The visualizations of the numerical simulations give further insight into the dynamics of the initial transient of the organ pipe. The data of the numerical simulations,

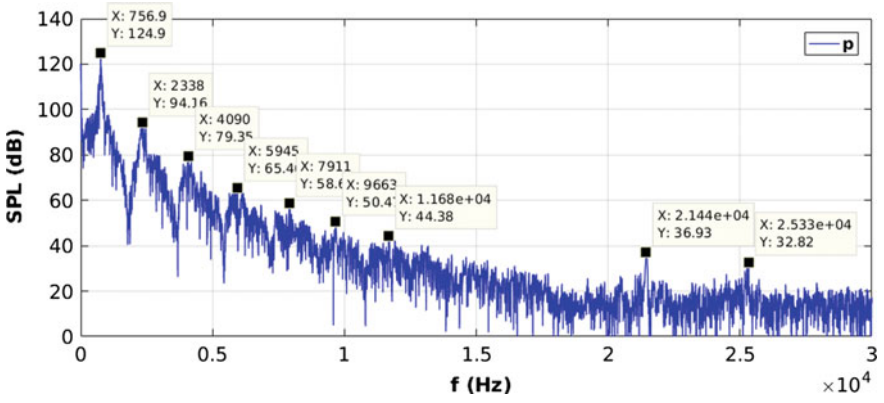


Fig. 3 SPL-spectrum of the time series at cross-section *cs_resonator*. The time series is generated by spatial integrating the data of the scalar field quantity pressure of the cross-section for the simulation time interval $t = 0$–100 ms. Seen are the fundamental frequency at $f_1 \approx 757$ Hz and the next odd harmonics $f_3 = 2338$ Hz, $f_5 = 4090$ Hz, $f_7 = 5945$ Hz, $f_9 = 7911$ Hz, $f_11 = 9663$ Hz, $f_13 =$11,680 Hz as well as two peaks at very high frequencies of 2.14 and 2.533 kHz

the spatial and temporal development of the physical quantities pressure, turbulent kinetic energy, vorticity and velocity magnitude are animated and provided as supplementary material.

The Fig. 4a–4i, 5a–5i, 6a–6i, 7a–7i and 8a–8i show snapshots of the development of the pressure in the time interval $t = 0.01$ ms to $t = 0.91$ ms which includes the first back-and-forth propagations of the initial pressure wave fronts along the longitudinal axis of the resonator. The pressure field is color coded in the range of 1011 hPa (light blue) to 1018 hPa (dark red).

One can see an emerging initial pressure fluctuation of high amplitude (red) generated by the organ pipe's jet flow, that enters the stagnant air column in the resonator. At the cut-up as well as at the wedge, the lower end of the labium, acoustical diffraction can be observed. This observation shows the excellent quality and the high level of precision of the numerical simulation performed in this study. At the inner walls of the resonator the initial pressure wave front is reflected. These reflections generate secondary pressure wave fronts, which propagate with higher velocities than the primary one (but on a longer way). This leads to an accumulation of the wave fronts. The propagation velocities of all observed initial pressure wave fronts are higher than the local speed of sound, they are supersonic! This indicates that the observed pressure waves have shock wave characteristics. The detailed analysis of the propagation velocities of pressure wave fronts will confirm this statement. The results are discussed in Sect. 3.4.

The initial flow produces two vortices (two dark blue dots side by side at the windway) that propagate into the mouth region, cf. Fig. 4f and the following. The vortex rotations, clockwise at the resonators side, anti-clockwise at the outer side create a dynamic pressure which is lowest in the vortex cores in accordance with Bernoulli's Principle.

The induced pressure wave front propagates spherically (circular in the 2D set-up) into mouth region of the organ pipe. The accumulation of the reflected pressure wave fronts at one hand and their attenuation on the other hand lead to an dynamical equilibrium that maintain the amplitude and the velocity of the primary pressure wave front. The primary pressure wave front therefore has properties of an self-reinforcing wave package. Waves like this are known as solitons. The dynamics observed here leads to a local increase of amplitude density in the upper end of the resonator at time t = 0.29 ms, cf. Fig 5f. While amplitude maximum of the primary pressure wave front is conserved its basis gets spatially distributed in the slipstream. This is shown in detail in Fig. 15. The superposition of the reflected wave fronts at the upper end of the resonator leads to a doubling of the number of traveling pressure wave fronts in the resonator's air volume, cf. Figs. 7a–i, 16 and 18a, b.

At the lower end of the resonator the primary pressure wave front reflects again. The high values of amplitude of the accumulated wave front basis induce high values of particle velocity. This can be interpreted as a sound source in the generator region. This is an important fact regarding to the formation of the sound field which will be discussed in more detail in Sect. 3.4. A fraction of the accumulated pressure wave escapes laterally through the mouth and radiates as a spherical sound wave into the free space. The main fraction of the pressure wave however propagates back in the

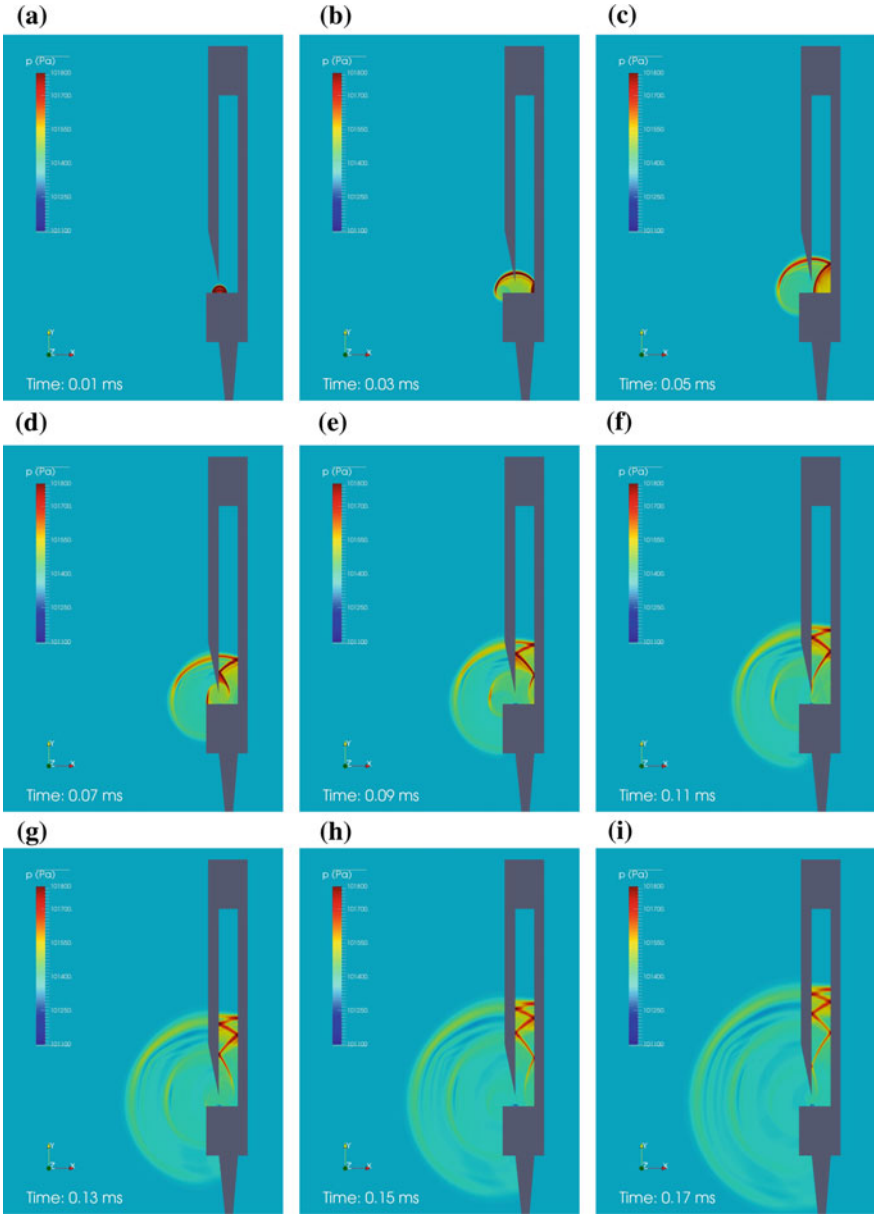


Fig. 4 Visualization of the numerical simulation. Color coded is the pressure at interval 1011–1018 hPa. The snapshots show the time interval 0.01–0.17 ms. Animations of the temporal and spatial development of the physical quantity are provided as supplementary material

resonator. Again, the reflections cause a doubling of the number of wave fronts, cf. Figs. 17 and 18c.

The generation of the initial pressure waves that show shock wave characteristics is initially driven by the jet flow. The jet flow works on a much slower time scale ($U/c_{shocks} \approx 1:20$) than the named pressure wave fronts, which are supersonic waves as pointed out. The driving process is interrupted by the pressure field occurring in the slipstream of the primary pressure wave front who disturbs the jet laterally by its produced particle velocity. This leads to an outward deflection of the jet. The jet's lateral sensitivity has its own time scale, determined by particle velocity, of the lateral acting sound pressure, the duration of the action and the inherent physical parameters of the jet itself, e.g. its flow velocity, its lateral thickness, its free propagation length, etc.

The jet flow changes its direction to a path outside the organ pipe, along the upper labium. With this change the jet flow locks up the pressure field of the resonator's air volume from that of the free space. The observations in Figs. 7a–i and 8a–i suggest that the lock up is provided by the inherent dynamic pressure of the jet flow. The jet responds to the fast impact of the pressure waves basis with a delay. While the jet is displaced and the primary pressure wave front propagates back to the closed end, dissipation processes generate a sonic pressure field in the slipstream of the primary pressure wave front. This is shown in Fig. 7d–i. The generated sound field in the resonator act on the jet. The pressure gradient between the free space and the air volume in the resonator supports the re-entering of the jet into the resonator, such that the jet continues with triggering the air column's oscillation.

The visualization has delivered first clues to solve the open questions regarding to the complex mutual interaction between the flow field of the jet and its self-generated sound field in the resonator which is induced by dissipative initial pressure wave fronts that have shock wave characteristics. In the following the dynamics of the pressure wave fronts propagating in the resonator in the initial transient will be examined in more detail.

3.4 Shock Wave Characteristics in the Initial Transient

To study the dynamics in the initial transient in more detail the propagation properties of the initial pressure wave fronts in the resonator are of special interest. The time development of the pressure field at cross-section *cs_resonator* is depicted in Fig. 9. Each time step is separated by an offset of 500 Pa for the sake of clarity. The plot shows the propagation of the primary pressure wave front followed by secondary pressure wave fronts which are caused by reflections at the resonator's inner walls. The velocity of the primary pressure wave front is constant. The velocities of the secondary pressure wave fronts change, but they remain higher than the primary pressure wave front's velocity. Figure 11 shows the results of the analysis of the velocities of the pressure wave fronts observed. Matter of interest were the propagation velocities of the amplitude maxima of the pressure waves fronts. I want to give two arguments

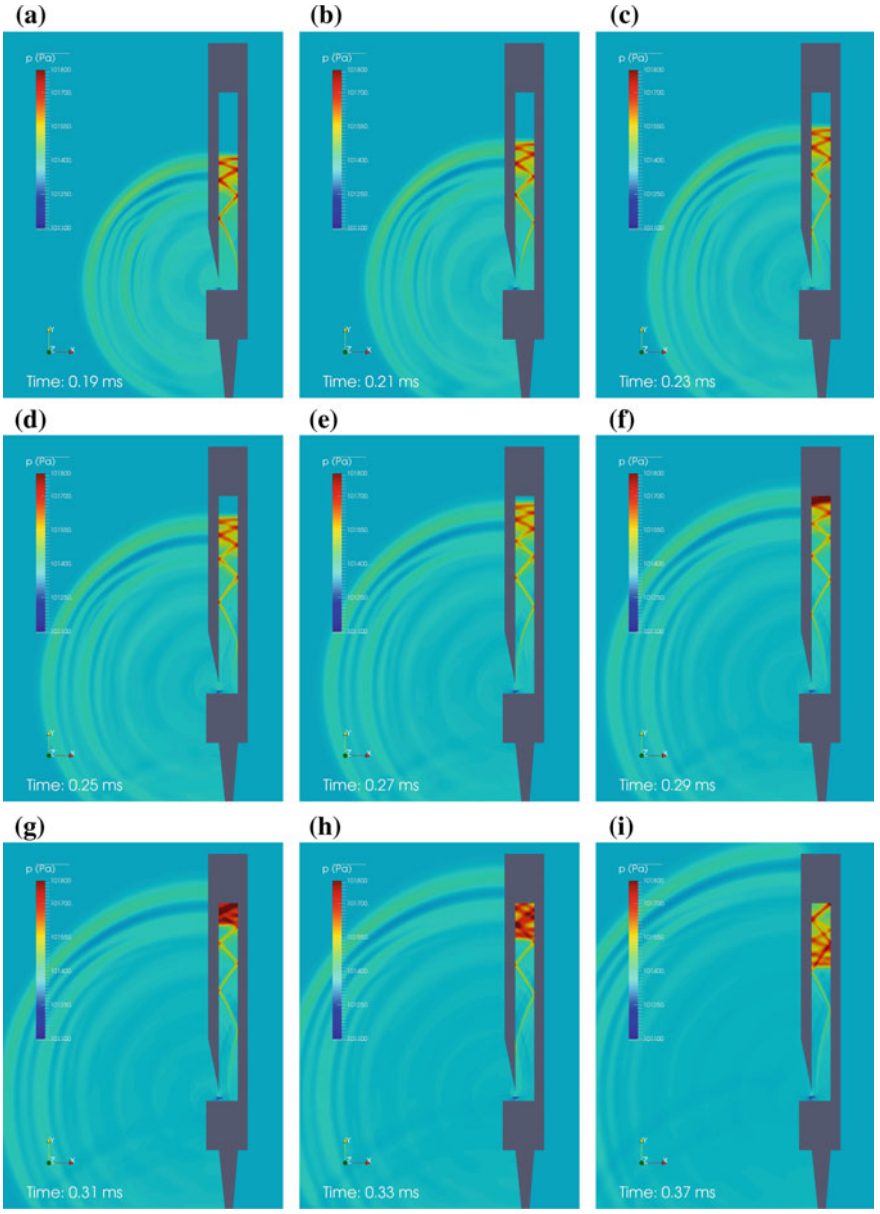


Fig. 5 Visualization of the numerical simulation. Color coded is the pressure at interval 1011–1018 hPa. The snapshots show the time interval 0.19–0.37 ms

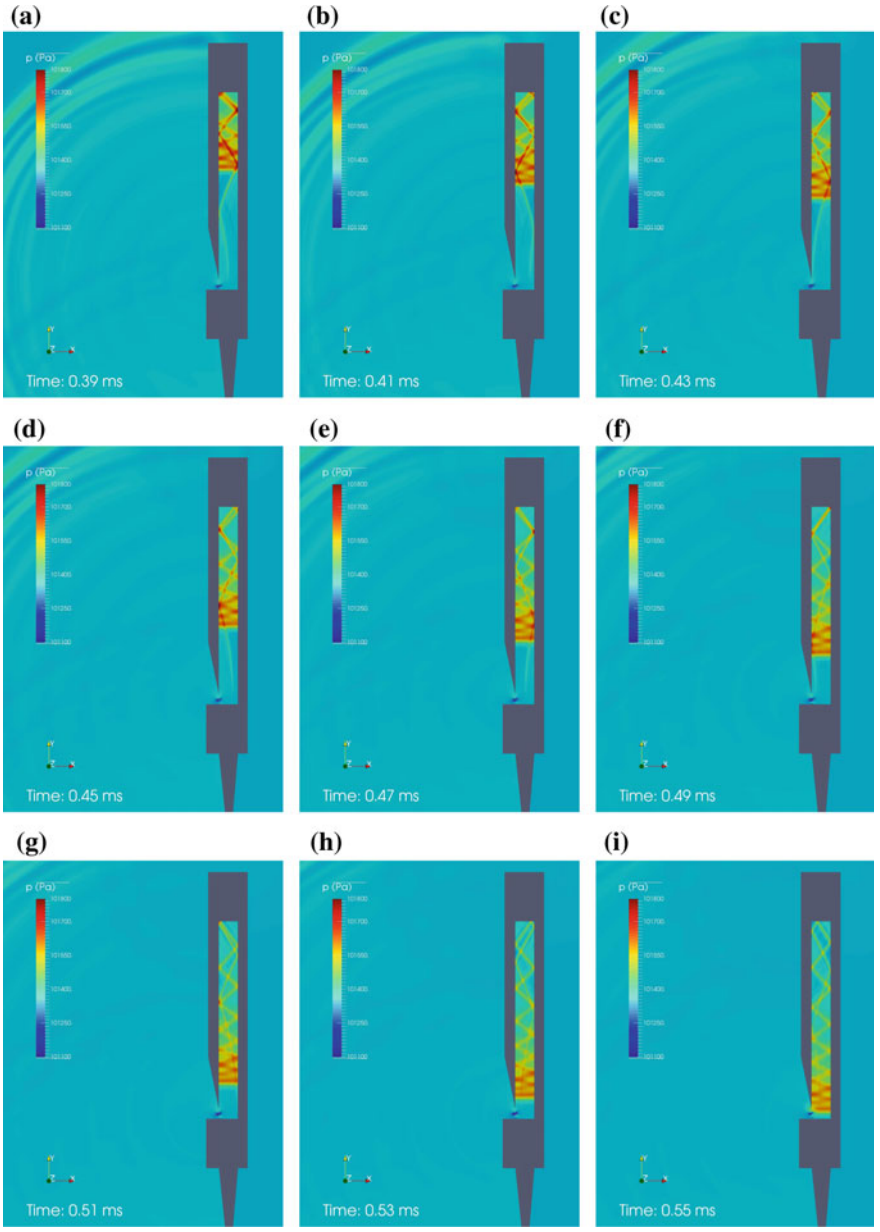


Fig. 6 Visualization of the numerical simulation. Color coded is the pressure at interval 1011–1018 hPa. The snapshots show the time interval 0.39–0.55 ms

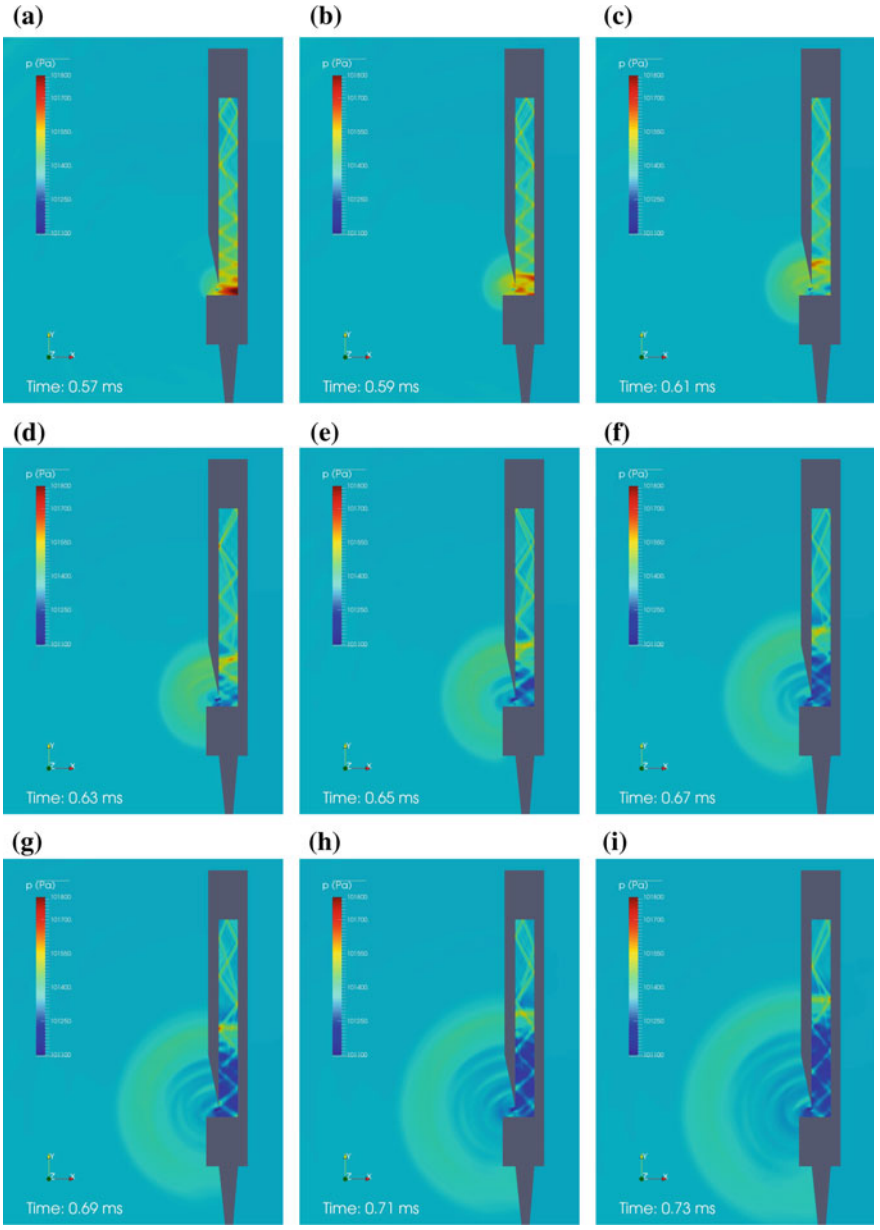


Fig. 7 Visualization of the numerical simulation. Color coded is the pressure at interval 1011–1018 hPa. The snapshots show the time interval 0.57–0.73 ms

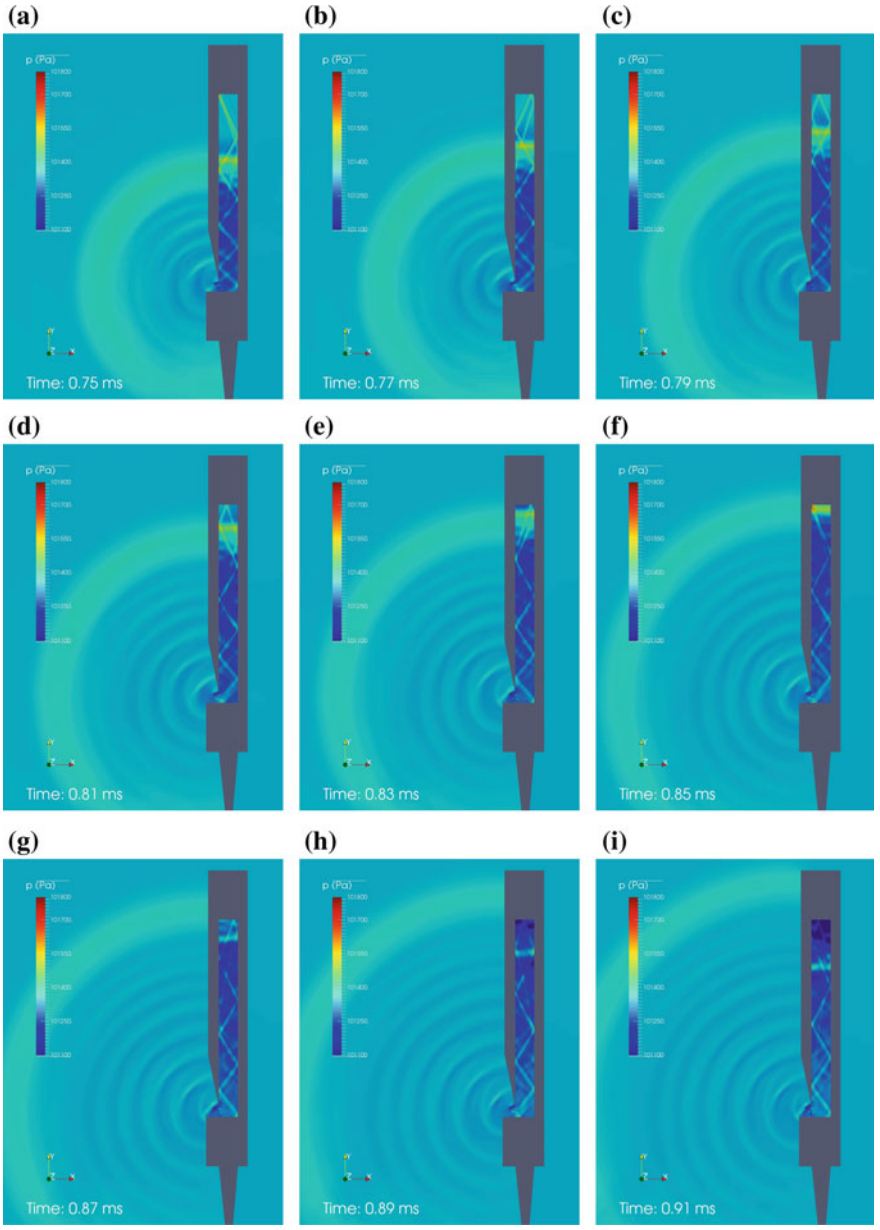


Fig. 8 Visualization of the numerical simulation. Color coded is the pressure at interval 1011–1018 hPa. The Snapshots show the time interval 0.75–0.91 ms

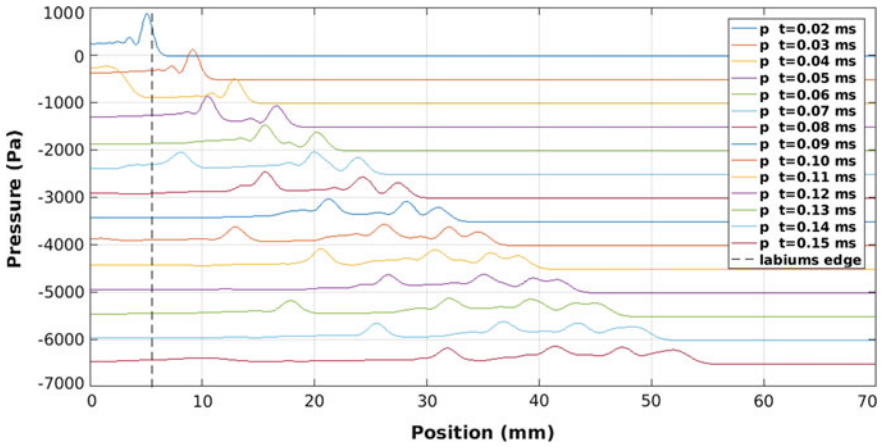


Fig. 9 Propagation of the initial pressure wave fronts along the longitudinal axis of the resonator. Depicted are the data sampled at cross-section *cs_resonator*. The time steps are separated by an offset of 500 Pa. The dashed line marks the position of the upper labium's edge which determines the height of the cut-up of the organ pipe of 5.5 mm

why for analysis the peaks of the pressure wave fronts were used. The first argument is that the addressed pressure wave fronts are coherent fluid mechanical objects which can be localized, distinguished and therefore be tracked best by their main property, the amplitude's maximum. The second argument is directly linked to the physical quantity we ask for. Tracking the velocity of the peaks of the pressure wave fronts ensures that we lock for the upper limit of the velocity of the quantity of interest, the pressure fluctuation. Any dissipative property of the considered fluid mechanical coherent object would lead to slower velocities. Any accumulative property leads to a re-build of the primary amplitude maximum at the pressure wave front. This can be retraced by the reader by studying Fig. 12 where both the damping as well as the accumulation of the initial pressure wave fronts are shown, in particular the accumulation and the re-build of the primary pressure wave front at times $t = 0.13$ ms to $t = 0.15$ ms.

The velocity of the primary pressure wave front is $c_{1.shock} = 363$ m/s and thereby higher than the local speed of sound of $c_0 = 343$ m/s. This confirms the statement that the observed pressure wave front has shock wave characteristics. The mean velocities of the secondary pressure wave fronts are even faster with $c_{2.shock} = 408$ m/s and $c_{3.shock} = 457$ m/s, but without overtaking the primary pressure wave front, so that one identifies the secondary wave fronts also as pressure wave fronts with shock wave characteristics. The named wave fronts show accumulation properties as well as dispersion and attenuation properties, which are discussed in the following section. The numerical results are compared with measurements of the velocities of amplitude maxima of the initial pressure wave fronts that propagate in a real pipe. Therefore a pipe with a length of 500 mm was utilized. The pipe was connected to a recorder's mouthpiece. The measurements have been done using an equidistant distributed

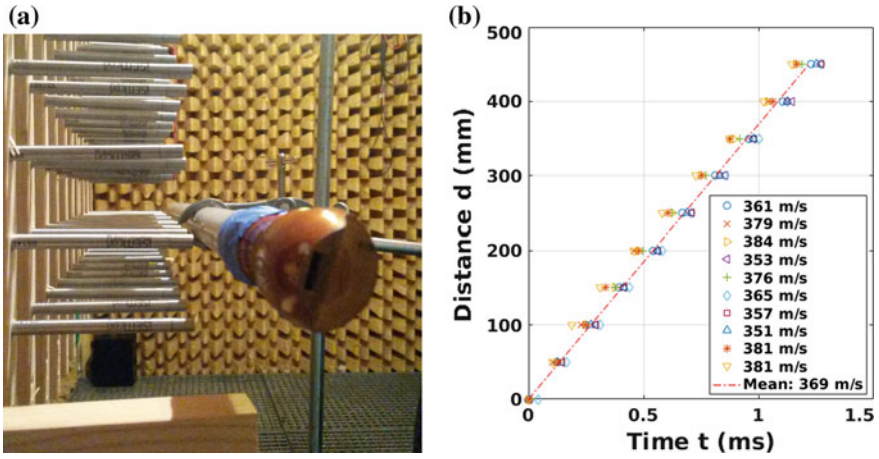


Fig. 10 **a** Experimental set-up to measure the initial pressure wave fronts propagating in a pipe. The pipe has a length of 500 mm. The mouthpiece is one of a recorder. The pipe was blown with the mouth. The measurement were repeated 10 times. **b** Depicted are the results of the measured speeds of the primary pressure wave fronts in the experiments. The dots are the data at each microphone. The mean velocity of all conducted measurements of the primary pressure wave fronts is marked by the red dash-dotted line. The value is $v_{1,shocks,e} = 369\,\text{m/s}$, which is in excellent accordance with the values found in the numerical simulations

microphone array line of 10 microphones (iSEMcon) in an anechoic chamber, cf. Fig. 10a. The measurements were done at the following conditions, temperature $T = 18.5\,°\text{C}$, humidity $L_h = 40.5\%$ and static pressure $p = 1016\,\text{hPa}$, which lead to a local speed of sound of $c_0 = 342.95\,\text{m/s}$. The sampling rate of the measurements was 44.1 kHz. The pipe was blown with the mouth. The results of the experiment are depicted in Fig. 10b. Although the experimental set-up differs from that of the numerical simulation the results are comparable, because of the fact that in both cases the point of interest is on the propagation velocities of the initial pressure wave front maxima. The velocities of the primary pressure wave fronts measured in the experiment are between 357 and 381 m/s, cf. Fig. 10b. The averaged value taken over 10 measurements is 369 m/s. The results of the numerical simulations are in very good accordance with the experimental results.

3.5 Attenuation, Accumulation and Nonlinear Damping

Displayed in Fig. 9 is the propagation of the initial pressure wave front in the resonator of the simulated organ pipe. One recognizes a dispersion of the primary pressure wave front. The observed dispersion produces a significant spatial distributed basis slipstream of the pressure wave front. The spatially distributed basis indicates that parts of the shocks are damped to slower propagation velocities than the primary

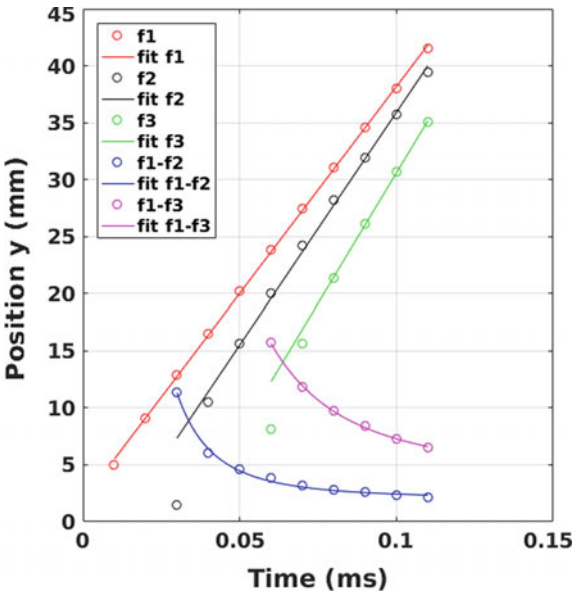


Fig. 11 Data from the numerical simulation. Depicted are the velocities of the peaks of the initial pressure wave fronts which propagate in the resonator of the simulated organ pipe in the initial transient. Marked by red circles is the propagation of the maximum of the primary pressure wave front along the cross-section *cs_resonator*. The red line fits the data. The slope gives the velocity of the peak of the primary pressure wave front which is $c_{1,shock} = 363\,\text{m/s}$. Labeled by black circles are the data of the second pressure wave front. Its linear fit gives $c_{2,shock} = 408\,\text{m/s}$. Not taken into account are the data points at the very beginning, where the velocity development is nonlinear. The data of the third pressure wave front are labeled green. The fit of the data gives $c_{3,shock} = 457\,\text{m/s}$. The circles and curves marked by the blue and the pink lines are the differences between the velocities of the secondary and the velocity of the primary wave front's maxima. In fact the secondary pressure wave front as well as the third one get damped in a nonlinear way relative to the primary one, but they are still fast enough to accumulate and rebuild the primary pressure wave front

pressure wave front has. On the other hand the accumulation of the secondary pressure wave fronts re-build the amplitude of the primary pressure wave front, cf. plot at $t = 0.15\,\text{ms}$ in Figs. 9, 15, 16 and 17.

The results of the analysis of the propagation velocities of peaks of the initial pressure wave fronts simulated by numerical simulations is shown in Fig. 11. One notices a nonlinear time development of the velocities of the secondary pressure wave fronts relative to the constant velocity of the primary pressure wave front peak at the beginning. Fits of the data give the following proportionalities: $(c_{1,shock} - c_{2,shock}) \sim t^{-0.365}$ and $(c_{1,shock} - c_{3,shock}) \sim t^{-2.518}$ of the damping. Although the damping, which is nonlinear, the velocities of the secondary wave fronts remain higher than the velocity of the primary wave front. This explains the accumulation of the wave fronts which leads to the maintenance of the primary pressure wave maximum. Responsible for the observed dispersion is internal friction. The supersonic velocities

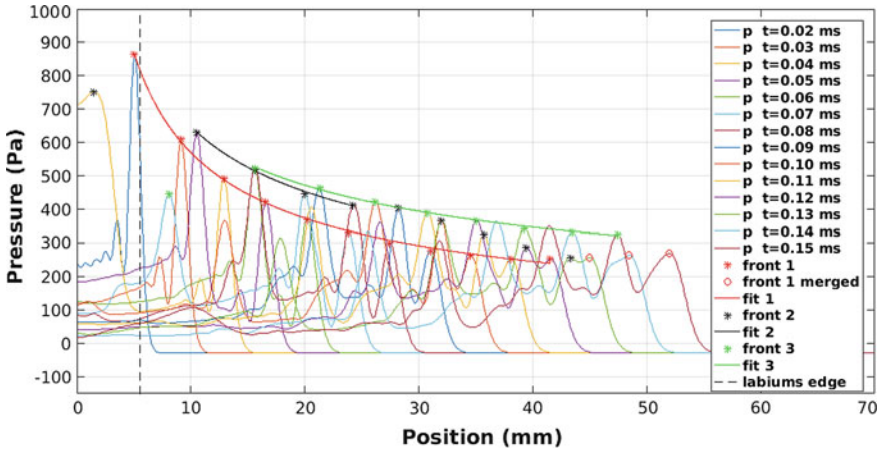


Fig. 12 Propagation of the initial pressure wave fronts along the longitudinal axis of the resonator. The data shown are sampled at cross-section *cs_resonator*

of the pressure wave fronts propagate against the inherent viscous properties of the medium, which indicates a high deceleration of the particles in the front. The deceleration leads to slower particle velocities ans less particle elongation. This is what energy transfer from supersonic into sonic time scales constitutes. In the absence of driving forces this process continues and energy is transfered into even slower times scales up to an complete dissipation [8] into the heat bath of the system. In the present case the primary pressure wave front is re-build by the accumulating secondary pressure wave fronts which propagate faster but dissipate faster too, such that an overtaking does not happen. The property of rebuilding a wave fronts amplitude by accumulative processes without an overtaking can be interpreted as a soliton characteristics. The details of this property are subject of the author's current research. Nevertheless the initial pressure wave fronts observed in the resonator are dissipative waves. With this results and statements an important aspect of the dynamics in the initial transient is quantified.

Another aspect is the attenuation of the amplitudes of the initial pressure wave fronts. The amplitudes of the pressure wave fronts observed in the numerical simulation attenuate in the initial transient. This is depicted in Fig. 12. Plotted is the time development of the initial pressure wave fronts. The superimposed representation of the data allows to calculate the corresponding attenuation constants for each pressure wave front at cross-section *cs_resonator*. The data are fitted by the following power-law functions

$$fit_{1.shock}(y) = 2314 \cdot y^{-0.6082} \tag{25}$$

$$fit_{2.shock}(y) = 2124 \cdot y^{-0.5154} \tag{26}$$

$$fit_{3.shock}(y) = 1778 \cdot y^{-0.4424} \tag{27}$$

One recognizes that the attenuation of amplitudes depends on the velocities of the peaks of the pressure wave fronts. The higher the velocity, the lower the attenuation exponent and therefore the fewer the attenuation. The amplitude attenuation is non-linear which is provided by the properties of the medium as well as of the properties of the resonator's walls, in particular their roughness. The wall properties of the organ pipe in the numerical simulation are modeled by an appropriate wall function, as already mentioned in Sect. 2.6.

3.6 Formation of the Sound Field in the Resonator

With extended visualization techniques more information about the dynamics in the initial transient are accessible. In this section the focus is on the formation of the sound field in the organ pipe's resonator. The time evolution of the pressure at cross-section *cs_resonator* is transferred into a color coded map depicted in Fig. 13. The time is on the x-axis and the length of the cross-section *cs_resonator* is on the y-axis with the lower end of the resonator at $y = 0\,\text{mm}$ and the closed end at $y = 100\,\text{mm}$. The introduced representation maps the pressure relative to the standard reference level of $p_{NHN} = 1013.25\,\text{hPa}$, which is color coded green and labeled as 0 Pa. Depicted is the time interval 0–20 ms. This is the relevant time interval for the dynamics in the initial transient which is of special interest in this study.

Figure 14a focuses on the first millisecond of the initial transient. The propagation of the primary pressure wave front can be identified as a sharp linear path, colored red. The secondary shocks occur as red dots because they just pass the cross-section at their zigzag way in the resonator, cf. Fig. 2. The primary pressure wave front reflects at the closed end of the resonator and propagates back with constant velocity. Note that the primary pressure wave front does not change its phase during the reflection. This is what is expected for pressure waves because of the boundary condition, the closed end of the resonator.

The lower end of the resonator is partly open at the mouth and therefore, from an acoustical point of view, one would expect a reflection of the sound waves associated with a change of phase of $\delta\varphi = \pi$, which would imply negative amplitudes of sound pressure. The observation is that the primary pressure wave front reflects at the lower

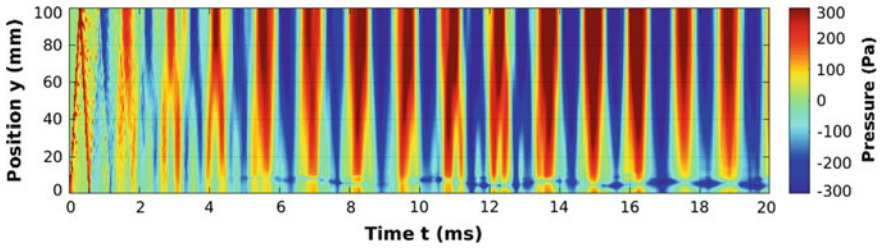


Fig. 13 Color coded time evolution of the pressure at cross-section *cs_resonator*

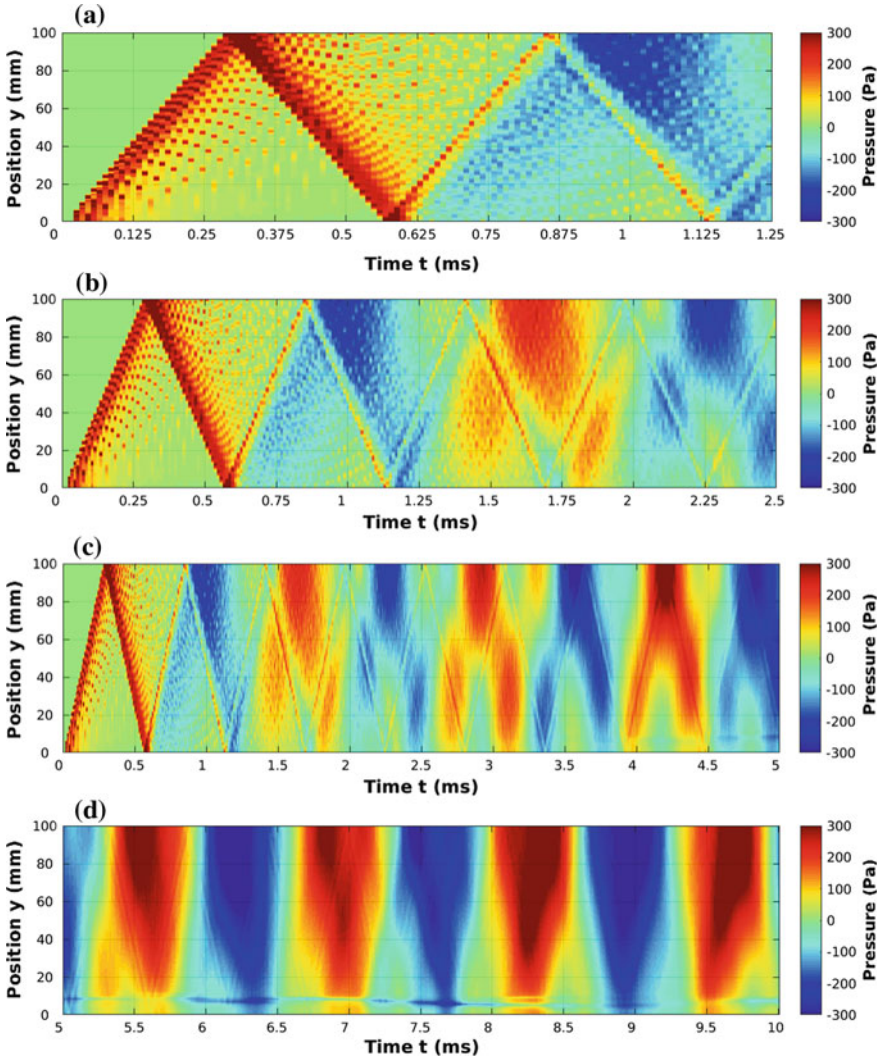


Fig. 14 Color coded representation of the development of the pressure field in the resonator at cross-section *cs_resonator*. **a** Time interval $t = 0$–1.25 ms. **b** Time interval $t = 0$–2.5 ms. **c** Time interval $t = 0$–5.0 ms. **d** Time interval $t = 5.0$–10.0 ms

end of the resonator not changing its phase nor its absolute value of constant velocity! That indicates that the primary pressure wave front does not see the lower end as an open end, but as a closed end. After the reflection the primary shock wave maintains its impulse-like characteristics as displayed in Fig. 14b–d. Also the secondary shocks reflect without phase change at the lower end of the resonator. Because of their higher propagation velocities ($c_{2.,3.shock} > c_{1.shock}$) they accrue and reinforce the primary pressure wave front (Figs. 15, 16 and 17).

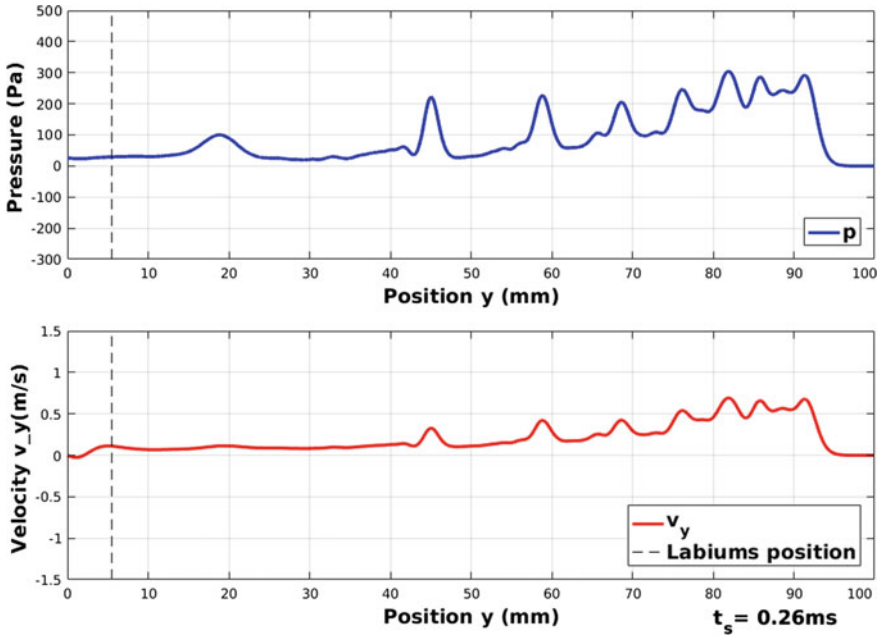


Fig. 15 Top: propagation of the primary pressure wave front and the reflected secondary pressure wave fronts along the longitudinal axis of the resonator. Bottom: the corresponding particle velocity. The data shown are sampled at cross-section *cs_resonator*. Animations of the temporal and spatial development the physical quantities are provided as supplementary material

However, one recognizes a pressure field with negative values relative to the reference pressure occurring behind the back-propagating primary pressure wave front, cf. Fig. 18c, d. The observation indicates that parts of the waves that travel in the resonator indeed reflect with a phase change of $\delta\varphi = \pi$ and 'recognize' the open end of the mouth. That implies that these fractions of the pressure field have sound pressure characteristics.

To explain this, one has to focus on the time scales of the different fluid mechanical objects that constitute the pressure field in the resonator. The fractions of the pressure field acting on the fastest time scale are the supersonic coherent objects ($c_{shocks} > c_0$). Their propagation velocities have impulse-like characteristics that means a directivity. They dissipate quickly if no re-build processes take place, because they work against the compressibility properties of the medium. A sound field acts on a next slower time scale ($c_0 = 343\,\mathrm{m/s}$). The speed of sound depends on temperature, density, humidity of the media. The jet of the organ pipe acts on an even slower time scale ($v_{jet} < c_0$). The time scale of the jet's flow is mainly determined by inertial and viscous forces. But, as already pointed out in Sect. 3.3, the jet has also a lateral sensitivity due to acoustical disturbances.

The lower resonator region of an organ pipe is bounded by walls and by the upper labium. It is open at the mouth's side, cf. Fig. 1a, b. The height of the cut-up

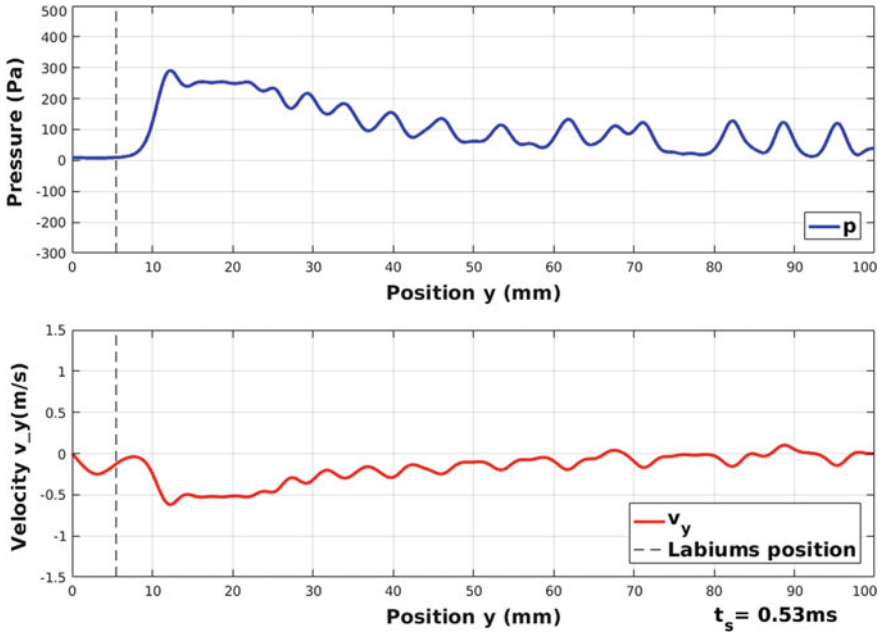


Fig. 16 Top: back-propagation of the primary pressure wave front and the reflected secondary pressure wave fronts along the longitudinal axis of the resonator. Note the doubling of the number of wave fronts by reflection due to the reflection. Bottom: the corresponding particle velocity. The data shown are sampled at cross-section *cs_resonator*

determines the free propagation length of the jet and therefore its fluid dynamical stability. The longer the free propagation length the larger the effects induced by the jet's inherent Kelvin-Helmholtz instability [3].

The observation is that the fast primary pressure wave front do not disturb the jet significantly. The reason for this is the shock wave characteristics, namely the direction of the k-vector of the wave front which is directed along the longitudinal axis of resonator and not lateral to the jet's flow, cf. Fig. 2. Whereas the spatially distributed basis in the slipstream of the primary pressure wave front is not supersonic but sonic because of the nonlinear damping manifest in the dispersion, cf. Sect. 3.5. Consequently, the basis in the slipstream of the primary pressure wave front has sound pressure characteristics and therefore it passes the jet and propagates through the cut-up into the free space. The observations in Figs. 6a and 7i reveal that the jet is disturbed by the sound wave produced by the slipstream of the primary pressure wave front. The reaction of the jet's flow is time-delayed because the flow acts on a slower time scale than the sound field does.

The distribution of the radiated sound wave is determined by the duration the spatially distributed basis in the slipstream of the primary pressure wave front acts on the jet. The radiation leads to a phase change of $\delta\varphi = \pi$ of the sound pressure in the resonator. The observed negative pressure, occurring in the slipstream of the

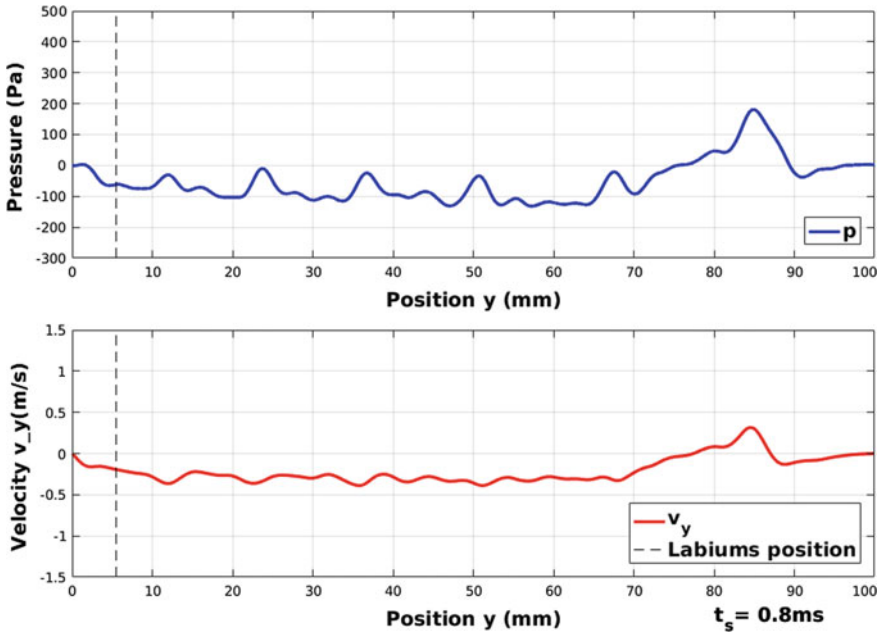


Fig. 17 Top: after the first reflection at the lower end of the resonator. Back-propagation of the primary pressure wave front to the closed end of the resonator. Development of a sound field (negative amplitudes relative to the reference pressure) in the slipstream of the primary pressure wave front. Bottom: the corresponding particle velocity. The data shown are sampled at cross-section *cs_resonator*

primary pressure wave front can be identified as the initial sound field in the resonator of the organ pipe. The described dynamics explains the origin of the sound field in the resonator.

The generated initial sound field is now triggered by the following processes: the periodically re-entering of the jet into the resonator and the energy transfer from the back-and-forth propagating pressure wave fronts into the resonator's sound field.

The jet's re-entering into the resonator is provided by the pressure gradient between the resonator's sound field and the free space. The assumption is, that the jet with its internal dynamic pressure separates the pressure field in the resonator from this in the free space. This can in principle also be traced back to the different time scales the jet and the sound field act on. The jet's lateral displacements are, despite of the fast impact by the sound field, determined by the jet's inherent stiffness, or inertia, which is determined by the flow characteristics. The jet is displaced by the sound field periodically but with a delay, caused by the different times scales as pointed out.

A stable sound can produced by the instrument if and only if the agents, the jet and the sound field in the resonator, adjust their oscillation properties to each other, namely their frequencies, their amplitudes and their phases. This indeed happens and the mechanism is discussed in the following.

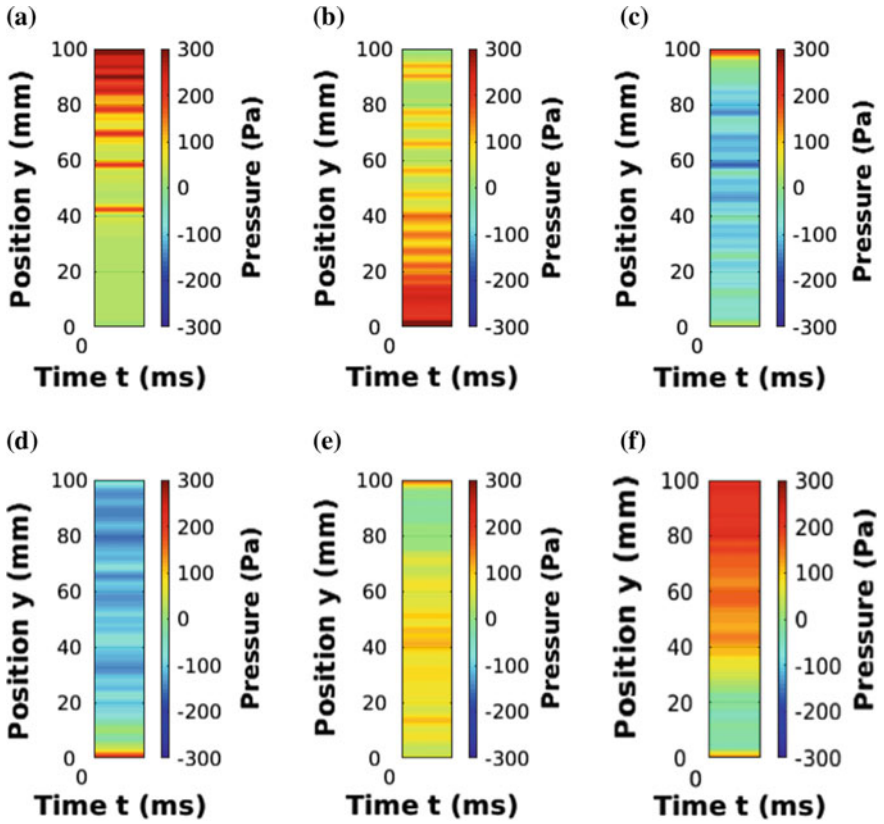


Fig. 18 The color coded pressure field at the cross-section *cs_resonator* at the first six reflection points located at $t_{R1} = 0.29\,\text{ms}$, $t_{R2} = 0.57\,\text{ms}$, $t_{R3} = 0.85\,\text{ms}$, $t_{R4} = 1.13\,\text{ms}$, $t_{R5} = 1.41\,\text{ms}$, $t_{R6} = 1.69\,\text{ms}$

In Fig. 14b–d the trigger process of the sound field can be studied. One recognizes the different time scales of the primary pressure wave front and the sound field. Also seen is the spatially distributed basis right behind the primary pressure wave front and the generated sound field after the reflections. In particular, note the increase of the sound pressure after the reflection at the closed end of the resonator at time $t \approx 1\,\text{ms}$, cf. Fig. 18c, d, and at later on at time $t \approx 1.7\,\text{ms}$, cf. Fig. 18e, f. This is a consequence of the dispersion of the pressure wave front during the propagation. As already mentioned the secondary pressure wave fronts are supersonic and faster than the primary pressure wave front but without overtaking the latter one. The dissipation of energy by damping (dispersion) into the fraction of the pressure field with the next slower time scale, which is the sound field. Whether the separation of the pressure regimes in front of and behind the primary pressure wave front, caused by the shock wave's characteristics, generates an additional depression or suction effect is not verified yet and subject of the author's current research.

The first re-entering of the jet into the resonator occurs at about $t = 0.9\,\text{ms}$. At this time the primary pressure wave front is already reflected at the closed end and it propagates back to the lower end of the resonator. Note, that the sound field initially starts its oscillation with a second harmonic characteristic, cf. $t = 1\,\text{ms}$ to $t = 4\,\text{ms}$. The sound pressure amplitudes increase at each reflection at the closed end. Subsequently the oscillation switches back to the fundamental frequency, cf. $t = 4\,\text{ms}$ to $t = 10\,\text{ms}$.

The increase of the amplitudes of the resonator's sound field is inter alia determined by the jet's mean flow velocity v_{jet}, which is related to the amount of energy per time interval, the power, that is transferred into the system by the jet, cf. Eq. (28). In the present case the velocity of the jet flow is not fast enough, which corresponds to an amount energy input that is not high enough to establish the second harmonics characteristic of the initial oscillations of the sound field in the resonator. Consequently the system switches to the energetically next lower state of possible frequencies, determined by the resonators length, which is the fundamental frequency. The slow down of the oscillations enables the system to increase its sound pressure amplitudes. These are triggered by the jet's re-entering. The mechanism is an important aspect of the interaction of the jet and the sound field. The amount of energy input by the jet causes the originally frequency of the sound field. In the present case this is the fundamental frequency of the resonator. The fundamental frequency acts on the jet and enslaves its displacements. In the case of a sufficiently high initial amount of energy input from the jet the system would switch into the next higher possible Eigenstate, which would be the third harmonic in the case of an organ pipe with a closed end. This is what happens if the instrument is overblown. Thereby the jet remains its lateral oscillation with the fundamental frequency, because its ability to oscillate is limited by the time scale of the jet flow.

$$P = \frac{\mathrm{d}}{\mathrm{d}t}\left(F \cdot ds\right) = \frac{1}{2} \cdot \dot{m} \cdot v_{jet}^2 + p_{static} \cdot \dot{V} + \dot{m} \cdot C_V \cdot T + \dot{m} \cdot g \cdot z \tag{28}$$

$$\text{with} \qquad \dot{V} = \frac{\mathrm{d}V}{\mathrm{d}t} = v_{jet} \cdot A \quad \text{the volumetric flow rate}$$

$$\text{and} \qquad \dot{m} = \rho \cdot \dot{V} \text{ the mass flow}$$

At time $t = 4.2\,\text{ms}$ the phases of the sound field and the primary pressure wave front coincide, which leads to a significant increase of sound pressure at the closed end. This indicates that the energy transfer into the sound field is particularly high at this time.

Following the theory of oscillations, an increase of amplitude (in the present case the amplitude of the sound field) cost more energy than the change of phase. This is caused by the asymptotic stability of the amplitude in opposite to the neutral stability of the phase [11]. The primary pressure wave front contacts the jet at $t = 1.11\,\text{ms}$. The lateral displacement of the jet can start again when the sound pressure in the slipstream of the primary pressure wave front reflects at the lower end. Thereby the sound field adjusts the phase relation of the jet's lateral displacement. This process is

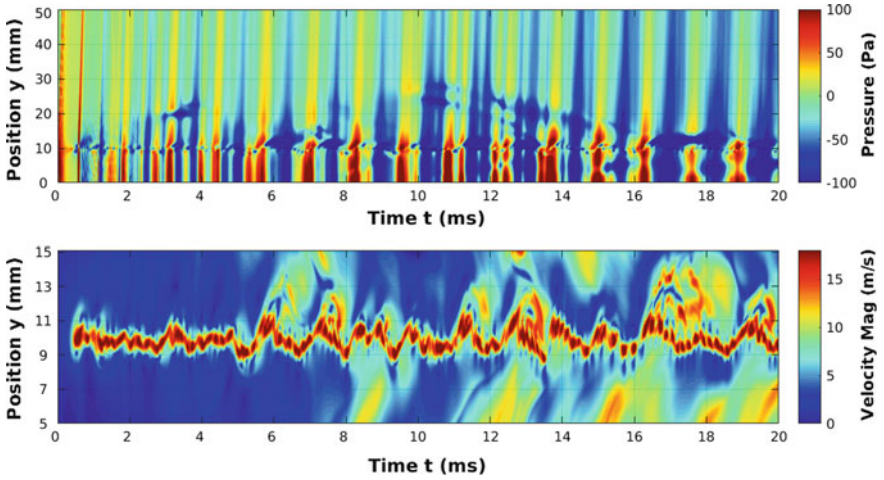


Fig. 19 Time development of the pressure field (top) and the velocity magnitude field (bottom) at cross-section *cs_jet*. The cross-section is located lateral through the jet at the distance of 3 mm from the windway, cf. Fig. 2. The cross-section has a length of 50 mm and 500 sample points. (Top) The interval $y = 0$–10 mm is the where the cross-section passes resonator, the interval $y = 10$–50 mm is the free space. (Bottom) The velocity magnitude field is color coded at the detail length 50–150 mm. The comparison of both plots illustrates the time-delayed reactions of the jet's flow to the impact of the sound field. The jet's oscillations are enslaved by the sound field after $t = 5$ ms

not done by a single step but piecewise during the next few reflections. The increase of sound pressure amplitudes in the resonator indicates a successful adjustment of the jet's displacements by the sound field. Because of the imperfect dynamics of the described process in a real, extended system, the adjustment process lead to a re-stimulation of higher harmonics. This constitutes the characteristics of the individual sound of the instrument, its spectrum of overtones that occur.

The next point of coincidence of the phases of the primary pressure wave front and the sound field is at $t = 8.9$ ms. Here the primary pressure wave front fully dissipates into the sound field. The assumption is, that the high amplitudes of sound pressure field in combination with the phase relation supports (maximizes) the dissipation process such that the primary pressure wave front completely decays.

The mutual interaction of the jet and the sound field of the resonator can be studied in Fig. 19, which illustrates the time development of the pressure field and the velocity magnitude field at the cross-section *cs_jet*. The location, the length and the resolution of the cross-section *cs_jet* is specified in Sect. 3.1. From about $t = 5$ ms the jet starts significant periodical displacements corresponding to the oscillations of the resonator's sound field. The jet's displacements and the oscillations of the sound pressure field in the resonator establish a phase relation which can be roughly estimated as $\varphi_{jet} - \varphi_{sound} \approx \pi$. Maximal displacements of the jet into the resonator's air volume correspond to sound field's amplitude maxima in the lower resonator region and vice versa. The time-delayed lateral jet oscillations are enslaved by the sound

field's oscillations , the fundamental frequency of the resonator. A more detailed analysis of the adjustment of the phase relation between the jet and the sound field as well as the role of the occurring vortex in the lower resonator region is subject of the author's current research (Fig. 20).

The mechanisms described in this section lead to a self-adjustment of the oscillations of the sound field in the resonator such that a stable sound of the organ pipe is radiated. The set of mechanisms which are described lead to a system that can be interpreted as a self-sustained oscillating system. The concept of an organ pipe as a self-sustained, extended oscillator is used successfully, in particular to investigate synchronization phenomena of an organ pipe and a loudspeaker [1] as well as of the synchronization of two slightly detuned organ pipes [5].

3.7 Modified Set-up, Smooth Initial Conditions

In this section results of numerical simulations with a modified inlet geometry as well as with more realistic initial conditions are discussed. The modified inlet is determined by a new mesh which includes the inner geometry of the pipe's foot.

Taking into account the gas dynamics in the pipe's foot is an important step to validate the numerical results of the case discussed in the previous sections. Without going into the details, one observes that the dynamics of the pressure field and the velocity magnitude field in the resonator is not essentially different of these of the case without the pipe's foot. The reader is encouraged to compare the numerical results illustrated in Fig. 21 with these of Fig. 19.

In Fig. 20b a snapshot of the case with pipe's foot and a ramp as initial condition at the inlet is depicted. The corresponding color coded representation of the time development of the pressure field and the velocity magnitude field at cross-section *cs_jet* shown in Fig. 22 illustrates the more realistic initial transient case, namely a smooth initial transient process. The ramp of the velocity of the inlet is linear,

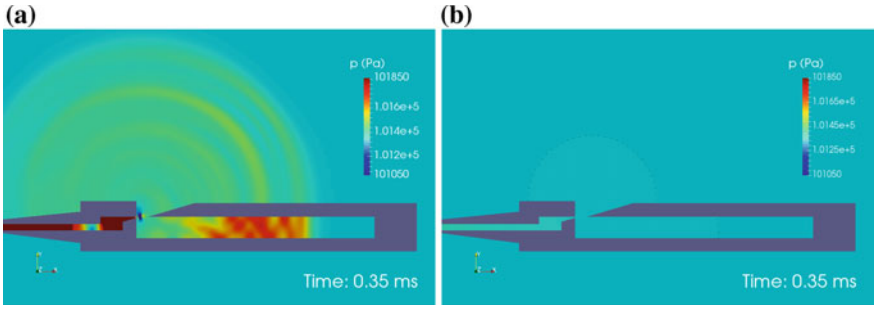


Fig. 20 Snapshots of the scenarios with calculated dynamics in the pipe's foot. **a** With no-ramp initial condition for the inlet and, **b** with a linear ramp as initial condition for the velocity of the inlet

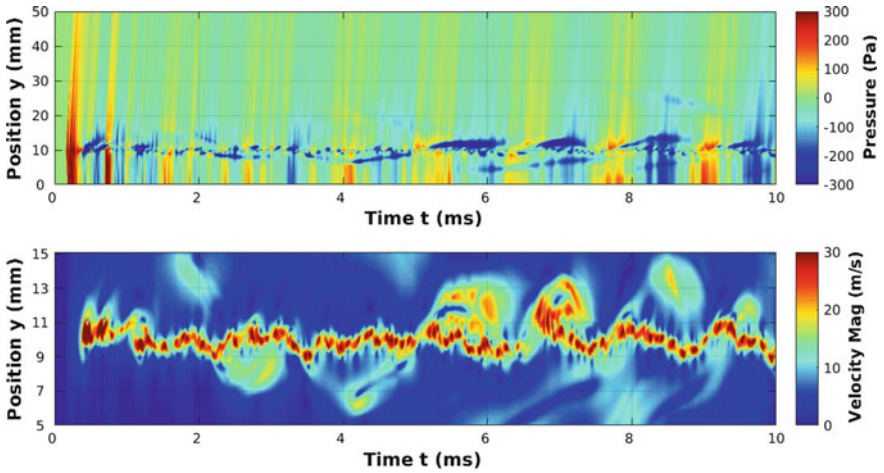


Fig. 21 Time development of the pressure field (top) and the velocity magnitude field (bottom) at cross-section *cs_jet* of the case with initial condition no-ramp but with the calculated dynamics in the pipe's foot

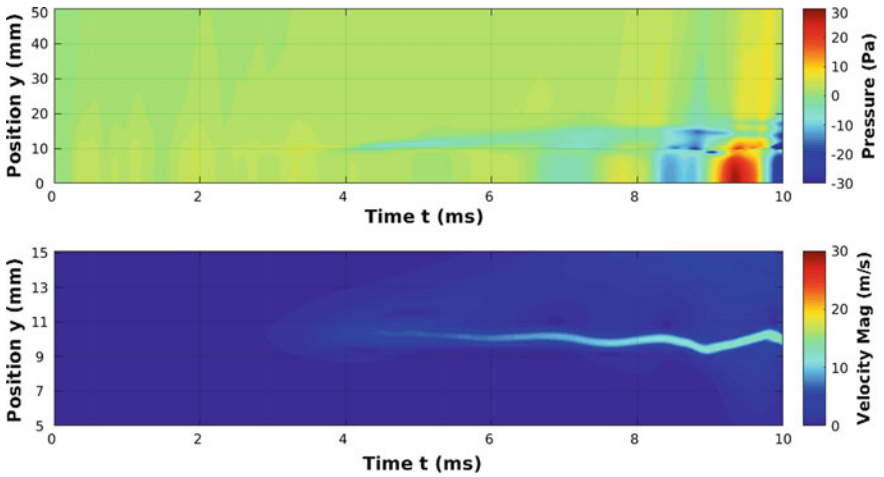


Fig. 22 Time development of the pressure field (top) and the velocity magnitude field (bottom) at cross-section *cs_jet* of the case with initial condition ramped velocity (0–5 m/s at time interval $t = 0$-0.02 ms) at the inlet. The scenario contains the calculated dynamics in the pipe's foot

increasing from $t = 0$ ms to $t = 0.02$ ms with values of $v = 0$ m/s to $v = 5$ m/s. Note that the initial velocity in the modified case is much lower than this in the unmodified case. The reason for that is that the tapered geometry of the chamber in the pipe's foot leads to a high compression of the air, which results in a comparable exit velocity at the windway as determined in the unmodified case, cf. Sect. 2.6. The time development of the modified case with the ramp as initial condition for the inlet velocity shows

increasing displacements of the jet starting from about $t = 4\,\text{ms}$. The reason for this displacements are initial pressure wave fronts of low amplitudes labeled with a black dotted line in Fig. 20b. That indicates that the mechanisms of sound field formation is, in principle, the same as discovered at the unmodified case. The detailed analysis and the comparison of the results flanked by more detailed measurements on real organ pipes and other wind instruments is subject of prospective publications.

3.8 Conclusion and Outlook

In this chapter the initial transient of a wooden stopped organ pipe is discussed. In the first section general notes on implementation and run of complex numerical simulations of aeroacoustical problems were pointed out. It made specific reference to the constitutive equations, the relevant fluid mechanical characteristic numbers, the Kolmogorov scales of the problem and the pursued grid size. Soft- and hardware decisions were addressed. Thermo-physical properties, boundary- and initial conditions, the utilized turbulence model and the mesh were stated.

In the second section the utilized methods of analysis and the results of the numerical case study were discussed. The data sampling were addressed and the spectral analysis of the cross-section *cs_resonator* were discussed. With advanced visualization techniques the complex dynamics of the initial transient process in the resonator were discovered. Thereby the occurrence of pressure wave fronts that have shock wave characteristics were identified. Important properties of the named pressure wave fronts were analyzed, in particular their velocities, their attenuation, their accumulation and their damping properties which are nonlinear. As a main result the mechanisms of sound field formation in the resonator caused by the initial pressure wave fronts and triggered by the jet's dynamics were discovered.

The complex dynamics of the resonator's pressure field in the initial transient is constituted by the mutual interaction of the jet flow, the initial pressure wave fronts, which are induced by the initial jet flow, the sound field, which emerges as a result of the dispersive characteristics of the back-and-forth propagating pressure wave fronts in the resonator and the jet's displacements that is controlled by the fundamental frequency of sound field. This complex feedback mechanism can be interpreted as a self-adjustment of the jet by agents the jet itself induces, namely the initial pressure wave fronts and the induced the sound field. All participants interact to each other on different time scales.

The supersonic characteristics of the initial pressure wave fronts lead to accumulation as well as to dispersion, attenuation and nonlinear damping. While the accumulation is responsible for the soliton-like back-and-forth propagation of the primary pressure wave front, dispersion, attenuation and nonlinear damping lead to dissipation of energy. Supersonic fractions of the pressure field, acting on fast time scales ($c_{shock} > c_0$) dissipate into sonic fractions, acting on the next slower time scale (speed of sound c_0), the sound field. The sound field is able to pass the jet and therefore to dissipate energy into the free space, namely the radiation of

sound waves. The sound pressure in the resonator for its part is able to disturb the jet laterally. The jet itself acts on the even slower time scale, characterized by the dynamic pressure ($v_{jet} > c_0$). Finally the jet dissipates energy via friction and generating vortices into the thermal energy reservoir, represented by the static pressure ($p = 101325\,\text{Pa}$, $v = 0\,\text{m/s}$).

Following the described cascade of dissipation energy, the mechanisms of sound field formation in the resonator are completely monitored. The role of the dispersive initial pressure wave fronts as well as the trigger process of the jet were pointed out. The phase relations between the named participants, the jet, which is a coherent turbulent fluid mechanical object, the resonator's sound field and the initial pressure wave fronts determine success or failure of the self-adjusting process in the initial transient of the organ pipe. In the winning case this leads to a stable sound generation and a characteristic sound radiation. The reader is encouraged to retrace the results by studying the animations in the supplementary material.

Open questions like the separation of the pressure regimes in front of and behind the shocks, the separation of the pressure regimes of the resonator and the free space by the jet and the role of the occurring vortex in the lower resonator region were mentioned. To discover the jet properties as well as to quantify the jets sensitivity regarding to lateral acoustic disturbances is also of particular interest and subject of the author's current research. The presented results eventually give a substantial contribution to improving the understanding of the first principles of the sound formation in the initial transient of an organ pipe, one of the most beautiful musical instruments.

References

1. Abel M, Ahnert K, Bergweiler B (2009) Synchronization of sound sources. Phys Rev Lett 103(114301)
2. Campbell DM (1999) Nonlinear dynamics of musical reed and brass wind instruments. Contemp Phys 40(6):415–431(17)
3. Chandrasekhar S (1961) Hydrodynamic and hydromagnetic stability. Oxford-Clarendon Press and New York-Oxford Univ, Press
4. Fabre B, Hirschberg A, Wijnands APJ (1996) Vortex shedding in steady oscillation of a flue organ pipe. Acustica Acta Acustica 82:863–877
5. Fischer JL (2014) Nichtlineare Kopplungsmechanismen akustischer Oszillatoren am Beispiel der Synchronisation von Orgelpfeifen. Ph.D. thesis, available at University of Potsdam
6. Fletcher NH (1976) Transients in the speech of organ flue pipes—a theoretical study. Acustica Acta Acustica 34(4):224–222(10)
7. Hirschberg A (1998) Shock waves in trombones. J Acoust Soc Amer 99, 1754 (1996). https://doi.org/10.1121/1.414698. South
8. Kolmogorov AN (1941) The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers. Dokl Akad Wiss USSR 30:301–305
9. Morse PM, Ingard KU (1968) Theoretical acoustics. Princeton University Press, Princeton, NJ
10. OpenFOAM^R—The Open Source Computational Fluid Dynamics (CFD) Toolbox Organization—OpenCFD Limited (2016). http://www.openfoam.com/
11. Pikovsky A, Rosenblum M, Kurths J (2001) Synchronization–a universal concept in nonlinear science. Springer, Berlin

12. Schlichting H, Gersten K (2003) Boundary-layer theory. Springer, Berlin
13. Alexander Schuke Orgelbau Potsdam GmbH (2018). http://www.schuke.com/
14. Verge MP, Fabre B, Mahu WEA, Hirschberg A, van Hassel RR, Wijnands APJ, de Vries JJ, Hogendoorn CJ (1994) Jet formation and jet fluctuations in a flue organ pipe. J Acoust Soc Amer 95:1119. https://doi.org/10.1121/1.408460
15. Yoshikawa S (2000) A pictorial analysis of jet and vortex behaviours during attack transients in organ pipe models. Acustica Acta Acustica 86(4):623–633(11)

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 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.

Computed Tomography as a Tool for Archiving Ethnomusicological Objects

Sebastian Kirsch

Abstract Musical instruments in ethnological collections can be a challenge for museums. Objects with uncertain provenance or doubtful circumstances of acquisition are considered to be repatriated. Some objects consist of sensitive material like human remains and are therefore bound to ethical guidelines for exhibition. On the example of a Tibetan *damaru*, a drum made of two human skulls, the provenance of the object and ethical considerations are discussed. For the case of repatriation 3D computed tomography is presented as a powerful examination and archiving method. Furthermore, virtual presentation and research concepts as well as other museum applications of 3D data of musical instruments are considered.

1 Introduction

In an ethnomusicological context, musical instruments can be regarded as unifying objects between tangible and intangible cultural heritage. In European museums they often function as representatives for a cultic group, people, a particular ritual or musical style and are subject to examination for a variety of purposes or intentions. In the cultural context of their origin, instruments are always part of a certain performance practice. The material of the instrument plays an important role as well as the purpose of the object and the connected musical occasion. Many objects in European ethnological collections were collected during the time of colonialism. Today, the circumstances under which the objects are acquired is reconsidered and in some cases a successful repatriation is the result of an intensive research on provenance. In the age of digitization 3D X-ray computed tomography (CT) can be a non-destructive method to archive objects in the case of repatriation. This powerful

S. Kirsch (✉)
Musical Instrument Museum University of Leipzig, Johannisplatz 5-11, 04103 Leipzig, Germany
e-mail: sebastian.kirsch@uni-leipzig.de

R. Bader (ed.), *Computational Phonogram Archiving*, Current Research in Systematic Musicology 5, https://doi.org/10.1007/978-3-030-02695-0_14

technique produces a 3-dimensional digital representation of an object which can be used for extensive research on the construction and materiality. This technology can also be used to produce virtual exhibitions or reproductions using 3D-printing or CNC-milling technologies. For example, a Tibetan drum made from two human skulls, a *damaru* in the collection of the Musical Instrument Museum of the University of Leipzig, was scanned using industrial X-ray CT during the MUSCIES-project. This instrument will serve as an example in the following discussion that concerns many ethnomusicological objects. Furthermore, the obtained data shows the capacities of CT digitization and possible applications including archiving.

2 The Object and Its Provenance

A *damaru* is an hourglass shaped drum used for various rituals in Buddhism and Hinduism. It is a right hand attribute of Shiva in his form as Nataraja, the King of Dance. It consists basically of two connected bowls or hemispherical bodies, connected at their top sides. Around this joint some decorative lace is wrapped, including two strings having balls made of wax and resin at their end. By turning the drum with one hand in alternate clockwise-and-counterclockwise movements these two whirling balls sway from one side to another. In the form how it appears in the Leipzig collection, it is connected to a grander religious context. The two bowls are made of the erania of human skulls and thus the object refers to be a Tibetan *Čang Teu*, which is used in tantric rituals, for example in Gcod practice [8, see, page 69]. The tantric *damaru* is traditionally "fashioned from the joined craniums of a fifteen or sixteen-year-old Brahmin boy and girl, or a sixteen-year old boy and a twelve-year-old girl" [2, pp. 107–108]. The male and female skulls also represent a "sounding together" and symbolize the union of method and wisdom as relative and absolute *bodhichitta*, which is the enlightened-mind. The fact that it is made of human remains is of major importance for the broader religious context. The skulls need to be taken from a charnel ground where the tradition of sky burials is practised. In this type of funeral the body is cut into pieces and placed on a mountain top to be exposed to the elements or eaten by animals. "Human bone and other necromantic or 'magical substances' (Tib. *thun*) are often explicitly prescribed for use in tantric rituals, because their properties endow the ritual implement or 'power object' with the specific affinities of the deity being propitiated." [2, p. 108] The *damaru*, for example, is often played together with a human-thighbone trumpet (*rkang gling*). This instrument is made from a bored human femur, preferably taken from a sixteen-year-old Brahmin girl in order to please the wrathful deities with its sound and to control spirit and elementals. Due to the loss of tradition and the extension of urban areas, charnel grounds are becoming rare and these objects achieve high prices. There are forgeries made from other recycled skeletons, but these are considered powerless (Fig. 1).

In the Gcod practice the topic of the ritual is connected to a transformation which can be understood as a small process of imitating death. "First, the participant views

Fig. 1 *Damaru* (MIMUL 2301) Musical Instrument Museum University of Leipzig, Photograph: Marion Wenzel

her own body as a symbolic representation of the universe to be offered as a gift to the Buddhas. Next, her own understanding becomes a ferocious goddess who dismembers her body with a knife and offers it as a feast for a host of fierce gods and demons. These are also attracted (and the meditation dramatized) by means of music." [8, p. 68] This disconnection of the bonds to misleading preoccupations or perceptions needs a dramatic synthesis of music, dance, meditation, and other ritual elements. The Gcod method dramatizes every philosophical teaching, every object and ritual action; expressing these elements both at surface and deeper meanings levels. So every component of the rite is highly charged with symbolism. The *damaru* itself, with all its possible decorations which can include engravings, gemstones, a decorated strap with sleigh bells and the hair of living and dead persons, can be considered as a microcosmic embodiment of the basic structure of the universe and of sentient life. The hollow body symbolizes the *dharmakaya*, which is one of the three bodies of Buddha. The skull drum is understood as balanced wisdom, its two drumheads as the union of appearance and emptiness. The decorative band around the central waist represents the thirty two major marks of the *sambhogakaya*, the body of enjoyment which is another important concept in Tibetan Buddhism. Furthermore, the strikers or a five-coloured silk valance tail refer to various religious ideas connected with the actual rite [2, see, page 108–108].

The understanding of Buddhist meditation methods and secular music has been a topic for research for many decades [6]. For a better understanding of the provenance of an instrument, the circumstances under which it was collected and the existing knowledge at that time have to be considered. The *damaru* in the Leipzig collection was first part of the collection of Peter Adolph Rudolph Ibach (1843–1892), founder of Rud. Ibach Sohn piano factory based in Wuppertal, Germany. He opened an

instrument museum in 1888 to the public. The *damaru* is found on a photograph of the Japanese room taken for the Festschrift for the celebration of the factory's centenary in 1894 [7] (Fig. 2). Due to missing documents on the acquisition of the instrument, it is not known how it was acquired by Ibach. In 1907 the Ibach collection was bought by Wilhelm Heyer and the *damaru* appears in a catalogue published in 1913 by Georg Kinsky. "No. 2299: Damaru, kleine Handtrommel mit beiderseitiger Bespannung; sehr altes Instrument, das heute nur noch von Schlangenbeschwörern und Affendresseuren benutzt wird." (*No. 2299: Damaru, small hand drum with covering on both sides; very old instrument, which is currently only used by snake charmers and monkey trainers*) [14, p. 220]. In 1926 the University of Leipzig bought the Heyer collection and since then the object has been preserved in its current location. It is remarkable that one of the founders of modern organology, Curt Sachs, mentions a *damaru* in his famous dictionary *Reallexikon der Musikinstrumente*. His description matches Kinsky's exactly, in fact word for word. Since both books were published in the same year, it can be assumed that either Kinsky took the description over as soon as Sachs' dictionary was published or Kinsky and Sachs had a correspondence about that topic. Sachs already notes that the skulls have to be taken from children's corpses. The proposition that the *damaru* is used by snake charmers and monkey trainers gives a short insight into the way eastern civilizations were perceived by the western world. Assuming Beer's remarks concerning the *damaru's* usage by "itinerant traders and street performers (…) since time immemorial to 'drum up' an audience with its staccato rattling sound evoking the urgency of a summons" [2, p. 107], it is possible to presume that this remarkable object made of human remains was attracting the attention of European colonialists not only due to its religious symbolism. It is possible that it was used as a sensational instrument in artificial performances.

The state of knowledge about Hinduism and Buddhism at the time of acquisition has to be considered when speculating on such circumstances. In William Rockhills *Notes on the Ethnology of Tibet*, published in 1895, there are two skull drums depicted, but little information is given about how to use them and there is nothing referring to the religious context of the instruments [22] . Only later, when Curt Sachs published his book on Indian and Indonesian music in 1915 a systematic analysis of Indian instruments in general and detailed information about the religious use of the *damaru* was available. He mentions that the cultic practice with the instrument is already very little and only active in the south of Tibet. "Im übrigen Land ist sie zu den Ausrufen, Bettlern und Schlangenbändigern herabgesunken." (*In the rest of the country it is degraded to the use of barkers, beggars and snake charmers*) [23, p. 76] For Francis W. Galpin the *damaru* reflected a primitive and barbaric culture. "Its very shape links it with primitive worship, for its barbaric form it was probably made of two human erania, like the Tibetan Čang Teu (Pl. X, 7) of the present day, taken from the skulls of slain enemies or of holy men." Due to its material it might have been a bizarre object for a collection that tried to illustrate the music of the world. However, as part of the Ibach collection and as it is described in the Kinsky catalogue, the collection story of the object informs about attitudes and approaches

Fig. 2 "Japanese Room" in the Ibach collection, the *damaru* (MIMUL 2310) at the wall on the left marked a with red circle, photograph taken from [7]

in early organology and the perception collectors had regarding the culture of origin of musical instruments and other objects.

3 Sensitive Objects and Codes of Ethics

The fact that the *damaru* consists of human remains is important for its use as a religious object. For a museum collection the presence of human material is definitely a challenge. According to the Code of Ethics published by the *International Council of Museums* (ICOM) "human remains and materials of sacred significance must be displayed in a manner consistent with professional standards and, where known, taking into account the interests and beliefs of members of the community, ethnic or religious groups from whom the objects originated. They must be presented with great tact and respect for the feelings of human dignity held by all peoples." [13, p. 25] In addition to the quite general guideline of ICOM, other organisations like the *Department for Culture, Media and Sport* in Great Britain or the *German Museumsbund* published recommendations for the care of human remains in museum collections [11, 20]. On the one hand, both publications show how the ethics of handling such sensitive objects is changing and respectful ways of storage and exhibition are found [10]. On the other hand, the need to study such objects is also expressed

by the "respect for the scientific value of human remains and for the benefits that scientific inquiry may produce for humanity" [11, p. 14] Therefore, museums are under a duty to provide access for visitors and researchers to the items as well as respecting cultural ethics and principles at once.

In this example, the discussion about this particular *damaru* can be considered from different perspectives. With the inclusion of parts of dead bodies a first thought in western culture could be to bury the object rather than to put them on display or in an exhibition. However, the *damaru* is a museum object and in case it shall be displayed it has to be treated with special respect. Apart from the recommendations cited above, legal issues concerning handling of human remains have to be taken in consideration as well. However, the way an object is presented in a museum has to satisfy ethical demands in respect to the cultural context of the object. To fulfil the demand of scientific research, the *damaru* can be the focus of many different scientific questions. As discussed before, the object was part of many different musical instrument collections for more than 120 years and has thus become an important historical source for examining the history of collecting. In a cultural context it can be compared to other objects dating from the same time or acquired at the same time from western collectors (comparison instruments: skull drum in the National Music Museum, Vermillion SD Inv. No. NMM 1383 or the *damaru* in the Victoria and Albert Museum Inv. No. IM.8-1932) and thus be included in broader research concerning collection and reception of tantric objects in the western world. Furthermore, the materials that the *damaru* is constructed of could also be used for research. An anatomist or an osteologist could possibly find out more about the provenance of the bones (Fig. 3). In this case the perimeter of the skulls and the texture of the fibrous joints seem to confirm the assumption that the skulls originate from children before the end of puberty. The origin and type of the skins used as drumheads can be determined, as well as the glue which was used to fasten them. The use of different textiles for wrapping around the waist could be part of the symbolic concept of the 'microcosm' *damaru* and it could be worth analysing them. If taking a sample for a DNA-analysis, it could be examined to determine if the bones really did come from a girl and a boy (Fig. 4).

The ICOM *Code of Ethics* describes that museums should respect the interests and beliefs of members of the community. The *damaru* is an object of use during meditation. By understanding the religious significance, this could validate its return to its original context or culture. This matches another point mentioned in the ICOM *Code of Ethics*. "Museums should be prepared to initiate dialogue for the return of cultural property to a country or people of origin." [13, p. 33] The history of repatriation of cultural heritage is still very short [see 21]. Some items like the Benin Bronze Statues are part of special exhibitions which address the topic of stolen art [16] but repatriation is rarely executed. Research projects on provenance often focus on the World Wars and seldom on Colonial history. In the United States of America the *Native American Graves Protection and Repatriation Act* (NAGPRA) has caused numerous cases of repatriation and seems to be a tool to reduce the illegal trade of indigenous items. In European ethnological collections, where objects from far abroad are hosted, the determination of possible legitimate owners or their

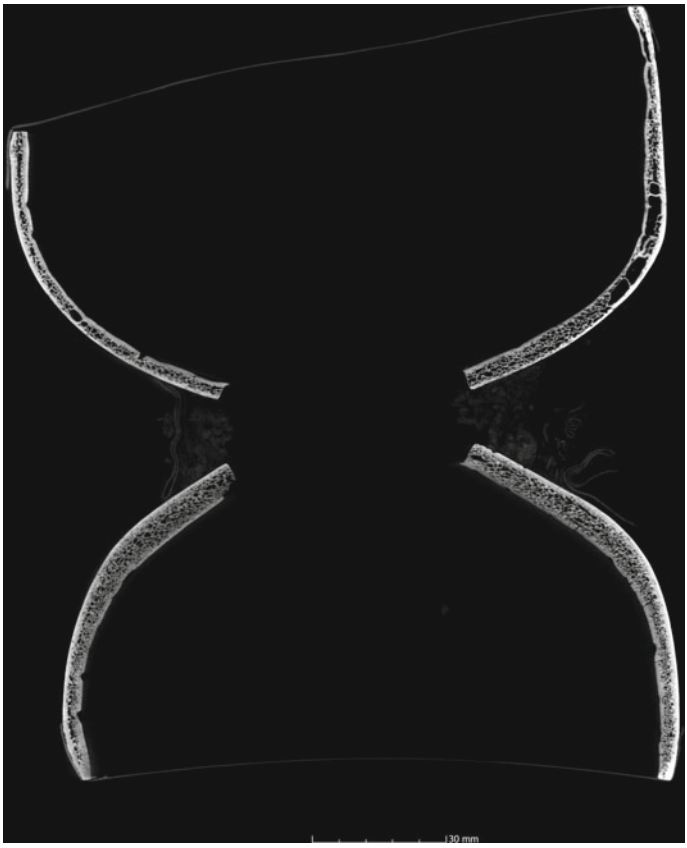


Fig. 3 Cross section of the *damaru*; details of the bone structure are visible

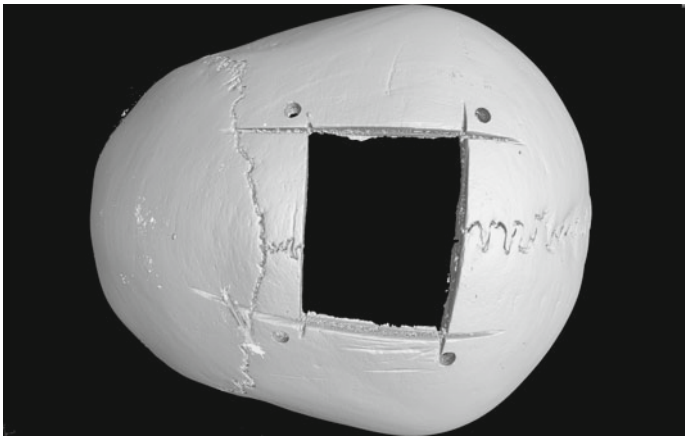


Fig. 4 Cutout and boreholes in one of the erania (MIMUL 2310)

ancestors is very difficult. However, repatriation can have many positive effects. "For some communities the repatriation of ceremonial materials from museums may be an important part of this process and linked to strategies to aid recovery from post-colonial trauma, and, as such, it has the capacity to contribute to indigenous health and well-being." [24, p. 122] Simpson points out that in "some indigenous communities the repatriation of human remains has also contributed to cultural renewal processes and stimulated the creation of new forms of contemporary cultural practices based on traditional values, ceremonies and art forms, thereby reinforcing cultural identity in the modern world." [24, p. 127] For the *damaru* the construction material (human skulls of teenagers buried on a charnel ground) is more and more difficult to acquire. Therefore, for this tradition or ceremonial life to be revived, dedicating the object again to the ritual context could be an act of cultural preservation.

4 Digitization as Archiving

Some institutions comply the demand of repatriation using digitization methods. The *National Museum of the American Indian* (NMAI) for instance has published archive material, such as photographs concerning indigenous people, online [5]. Countless museums like The *Ethnological Museum Berlin* have photographic images of their whole collection including musical instruments online. Some of the objects are additionally illustrated by a 3D PDF of a medical CT scan. Many museums are connected to the *Europeana* project or the *MIMO*-website (Musical Instrument Museums Online) [17]. The idea of making objects available is resulting in many virtual museums like the *British Museum* which can be partly explored with a 3D visit of some collections [19] or the *Virtual Museum of Canada* [25]. Many more similar projects are currently starting and the digital representation of collections will increase. The approach of such projects is to give digital access at a minimum to the collected material to anyone who is interested; in ethnological museums especially the community of origin might be addressed. In projects like the *Fourth's Museum* of the NMAI the team often consists of scientists with knowledge of indigenous background or the aboriginal peoples themselves are consulted to assure an ethically correct handling and accurately edited publishing of the objects. These modern methods to build digital archives of ethnological objects are not only deemed uncritical [4, 12] but are also considered to be a positive approach to make objects available, visible and researchable. Often such projects are connected with the investigation of material which has been stored in a depot for a long time, however, it is important to note that these methods do not give the actual objects back.

In the case that an object has to be returned or for the investigation of structural elements of the object, the most powerful digitization method for physical objects like ethnological musical instruments is 3D computed tomography.

4.1 3D X-ray Examination of Musical Instruments

3D computed tomography (CT) is an imaging method which is widely used for medical purposes and industrial non-destructive testing. During an X-ray CT scan numerous single images are recorded from different angles of an object. All single images are later computed to a 3D data set which is a digital representation of the object with all structural information about the inner and outer surface (apart from the colour) and internal parts. The object can be examined by cross sections in every direction, with surface renderings of all parts or by taking measurements at every (otherwise inaccessible) location and numerous further applications are possible. In a medical scanner the patient or the object lies on a cot and an X-ray tube and detector are rotating around him hidden in a white tube. The CT facilities for industrial applications are different. Here, the object is placed on a rotation table between X-ray source and detector. Unlike in the medical scanner, this setting is not optimized for human bodies but can be adjusted to the needs of different objects. Thus, higher energies can be used for the irradiation of objects with higher densities and better resolutions can be achieved. While the medical scanner is more or less fixed on a spatial resolution around 400 µm, industrial scanners can produce a higher quality where details in musical instruments from around 50 to 100 µm can be distinguished. During the DFG-funded *MUSICES*-project from 2014 to 2018 the *Germanisches Nationalmuseum*, Nuremberg scanned in collaboration with the *Fraunhofer EZRT* in Fürth more than 100 different musical instruments using various industrial and medical scanning methods in order to develop recommendations for the executions of such scans [15, 18]. The implications of this project are manifold. The developed workflow model can be used as an exemplar for an optimized and efficient scanning of a single instrument or a large amount of different objects. Due to the fact that musical instruments consist of many different materials and exist in every size, the guidelines can be transferred to other classes of objects. For instruments made of different materials with varying X-ray attenuation properties such as clarinets with keys etc., specific scanning methods using two different spectra were refined [26]. As far as image quality is concerned, it can be noted that a spatial resolution of 100 µm voxel size is sufficient for most purposes. Such a quality allows further examinations like dendrochronological dating of wooden parts and high precision measurement. Furthermore, the density of the irradiated material can be determined. Density is significant when performing acoustic modelling of musical instruments. The CT data can be used to determine density as well and thus the digitized virtual instruments can replicate accurate sound.

The *damaru* was also among the objects that were scanned during the *MUSICES*-project using an industrial CT scanner. The images give detailed information regarding the construction of the drum. For example, both hollows are connected with a square cutout. The surface rendering shows how the saw cut roughly in the upper area of the skulls. For joining both erania, four holes were drilled for a thread which binds both parts together. When focusing on the waist, it becomes obvious that the textile wrapping consists of different cords with distinctive windings (Fig. 5). Tak-

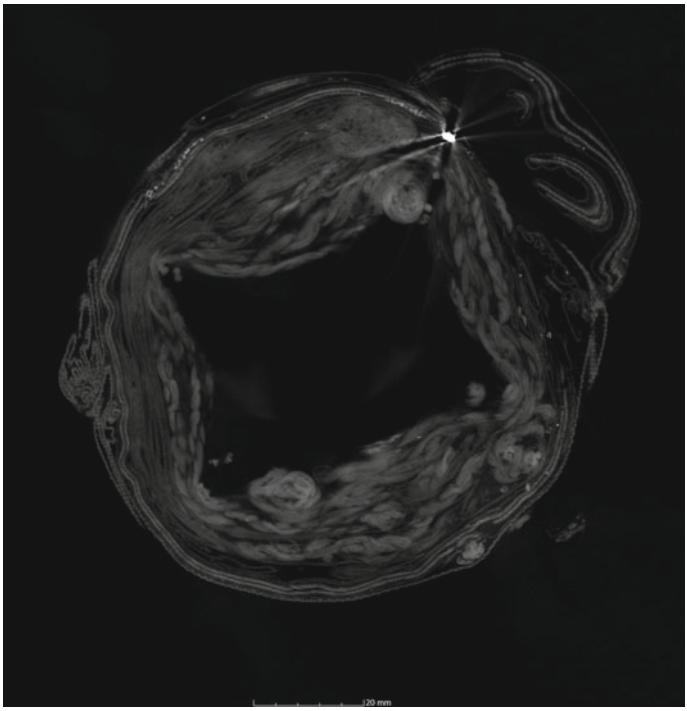


Fig. 5 Textile wrapping around the waist of the *damaru*, different layers are visible in the CT cross section

ing into account that all material of this object has a high symbolic meaning; the usage of different threads, even if they are not externally visible, could be considered part of the ritual context. The scans also show anatomic details. For example, the inner surfaces of the skulls have imprints of the anatomical structure of the brain (Fig. 6). Analysis of cross sections of the skull bones garners information about the internal structure of the erania. Advantages of using computed tomography such as the 3D-X-ray image include, the ability to gain non-destructive access to the internal workings of the object and the capacity to retrieve high quality, detailed information regarding the manufacturing technique and the internal physical appearance of the object. These insights can be a starting point to find out more about the construction of an instrument such as the density and the internal structure of the different parts. Information from the scans can also be used to detect potential repairs and damage, to compare with the internal structure of similar objects or as additional data to complement existing research. When applying segmentation, single parts can be extracted and compared separately. Furthermore, the scans can be a good source of information for the determination of authenticity or conservation and restoration measures. Using the X-ray CT data set enables research on the physical structure of the object and the required level of analysis for a museum object (as described in the guidelines cited above) to be executed.

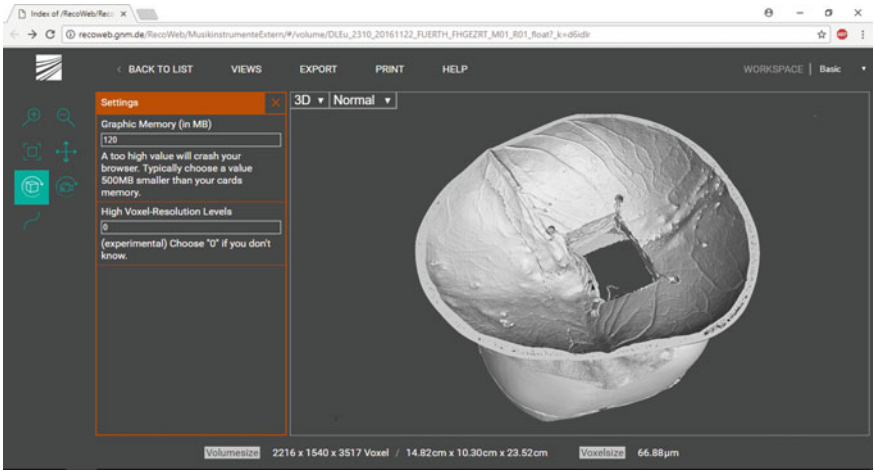


Fig. 6 The *damaru* (MIMUL 2310) displayed in the Fraunhofer RecoWeb viewer, accessible on the MUSICES homepage www.musices.gnm.de

4.2 Data Archiving and Accessibility

To ensure long term use and further examination for scientific purposes, a persistent data management system is critical. In the case of 3D CT of musical instruments, the *MUSICES* project developed a meta data model and database that can be used for analysis of the objects scanned during the project and will also be available for future projects and research. In this database, all relevant information on the object and the technical parameters of the scan are documented in detail.

A typical CT scan campaign can be grouped into three areas: the object description, the actual scan and the reconstruction and evaluation. Creating an object description prepares the CT scan and provides the required information for the scan. Size and material of the object as well as the way how it can be mounted on the rotation table determine the technical parameters and influence image quality. For the scan itself the technical set up (X-ray source and detector) and all parameters like tube voltage, current, measurement method, amount of single images etc. have to be recorded. In the third step of a CT examination process, the single image is converted to a 3D data set using specific algorithms, which is called reconstruction. The images can only be understood and replicable if this work step is also documented. At the end, the volume of images can be viewed and results can be evaluated.

In the *MUSICES* database all this information is correlated using a persistent identifier which consists of several elements like the scribal abbreviation of the collection (DLEu for the Leipzig collection according to the ICOM CIMCIM sigla list), the inventory number, the date, the place and the institution of the scan, the chronological number of the scan and the chronological number of the reconstruction. In the case of the *damaru*, the identifier and the name for the data set is:

DLEu_2310_20161122_FUERTH_FHGEZRT_M01_R01

The entire database is built using the CIDOC-CRM ontology (ISO 21127:2006), a reference model which provides definitions and a formal structure for describing the implicit and explicit concepts and relationships used in cultural heritage documentation. The database is more than a simple data storage system. The so called "WissKI" (German abbreviation for scientific communication infrastructure) organizes and connects the data sets according to the ontologically defined entering fields and allows a targeted workflow and organization of the different entries. All meta data is correlated to the actual volume data set in a bijective way.

The storage format has to be chosen carefully. Many providers of industrial CT scanner use proprietary formats and it can be questioned whether these formats will remain readable in the future. The format DICOM (Digital Imaging and Communications in Medicine) is used for many medical imaging techniques and presents a good choice for long term archiving. Given the law in many countries require the storage of medical imaging data for many years; the DICOM format has potential longevity.

One challenge for a database of CT scans is the online display of big data sets. Due to the high resolution of the scans, the volume data sets of objects are rather large. The data set of the *damaru* has a size of 21.6 GB. Compared to other musical instruments like a rabab, which was also scanned during the *MUSICES* project (size of data set: 69 GB) the data set of the *damaru* is quite small. Usually, these volume data sets can only be processed using hardware with high computational power and specialized software. For the *MUSICES* website a web viewer is used which can display big data sets via internet [9]. The entire 3D data set can be opened and analysed using only a common internet browser.

The database provides open access to research data without any required specialised hardware or software. This is possible due to a pre-processing step of the software which fragments the volume in small portions which are easier to display.

The *MUSICES*-database can be used as an exemplar for the combination of a persistent meta data archiving structure and the presentation of the actual volume data set. It hosts meta data and displays ca. 230 data sets of more than 100 musical instruments. This information can be connected to more specific databases, for example, on intangible cultural heritage. In conjunction with recordings or videos, the CT images of musical instruments could be a good addition to the representation of a musical practice and serve as a tool for archiving. As a consequence, open access to a rather large amount of data is provided to a huge community of researchers.

4.3 Applications

In case of repatriation the 3D data set can be used for further museums' purposes. Additional uses of the 3D data set include object repatriation by museums. Due to the high quality of the images, an exact replica can be produced using 3D printing technology. For this purpose the surface of the CT scan has to be determined and

extracted for example as a stl file. Current 3D printers achieve a very high precision and can process materials with different properties. For a reproduction of the *damaru*, a material that has similar properties and surface character like bones can be used. The other parts of the object could be remade using traditional techniques. Of course, such a copy can never replace the original object, but could serve as a substitute in case the object has to be repatriated. For the ritual act the fact that the instrument is made of human remains is of major importance. A plastic copy can tell a different story of collecting and usage and is also easier to display since it is not bound by the same ethical debate as human remains. In this context it is worth mentioning that the beginning of museum age the understanding of authenticity was different from the current view and the use of copies in museums for the illustration of certain topics was much more common. Plaster casts of ancient sculptures of bronze objects or entire columns were presented next to original objects.

In the digital age, the topic of authenticity is still a relevant discourse. Even if a 3D printed copy or the digital representation of an object cannot answer all scientific questions e.g. on the chemistry of the material, it can open new horizons for both research and exhibition. Objects are in their digital representation much more flexible and not bound to a museum's climatic conditions and in many cases free from ethical concerns. Apart from the scientific approach, the view inside the objects can be used for different education services in museums. Videos and pictures can complement the traditional audio guides, but also in virtual or augmented reality applications in the exhibition and beyond museums walls. Using applications like *Civilisations AR* published by the British Museum [1] objects can be virtually projected in every surrounding space using only a smart phone. The X-ray function allows moving through the 3D CT scan of an object. This expands the possibilities for communication concepts and addresses the demands of a digital society.

5 Conclusion—Building a Digital Heritage

Museums have the responsibility to preserve, disseminate and impart cultural knowledge as well as provide items for research. This concerns the cultural history of the object, provenance, material, construction and much more. In the case of human remains or objects which were acquired under doubtful circumstances, museums try to find suitable ways to satisfy ethical needs. 3D digitization methods like X-ray CT can influence traditional museum research, archival and communication methods. They can change the availability and access of objects and affect the way knowledge is created. They could provide a new, modern and progressive approach for handling repatriation and the treatment of sensitive objects. In the case of the *damaru*, the CT scan offers a lot of information on the physical appearance of the object but not on the chemistry or the aura in the sense of Walter Benjamin. If the object is repatriated its purpose as a museum object changes again into a cult object with a very specific ritual function and it could be part of the revitalization of a tradition. The *damaru* can serve here only as an example that has different representative features that concern

material and cultural context. In a similar way the discussion on the *damaru* pertains to many other objects in ethnomusicological collections.

The digital representation of an object has to be seen as something innovative. The 3D model is an artefact by itself and a product of a cultural progress. It has its own value and has different properties than the original. Thus, it is more than just a copy. It can forge new interactions with the original and a new relationship to museum objects. The UNESCO *Charter on Digital Heritage* published in 2002 claims an equally professional preservation for cultural heritage data as well as for physical objects and demands open accessibility to the generated data [3]. By using open formats and as well as through the integration of international databases like *MIMO*, *Europeana* and *MUSICES*, the objects can be connected in diverse research contexts or educational concepts. Modern database structures are capable to combine historic, social or humanistic research on the tradition and use of an object with scientific results on material analyses and a virtual representation. The digital model of the object can be preserved, explored and distributed even if the original object is repatriated.

Acknowledgements I gratefully acknowledge Prof. Dr. Ingo Bechmann, Institute of Anatomy at the University of Leipzig and Dr. Carsten Babian, Institute of Forensic Medicine at the University of Leipzig for important help with the examination of the skulls.

References

1. BBC (2018) BBC launches augmented reality app for Civilisations (cit. on p. 14)
2. Beer R (2003) The handbook of Tibetan Buddhist symbols, 1st edn. Shambhala, Boston, Mass 2003 (cit. on pp. 2, 4, 5)
3. Charter on the preservation of digital heritage (2002) (cit. on p. 15)
4. Christen K (2009) Access and accountability. Anthropol News 50(4):3–5 (cit. on p. 8)
5. Crouch M (2010) Digitization as repatriation. J Inf Ethics 19(1):45–56 (cit. on p. 7)
6. Cupchik JW (2013) The Tibetan gCod Damaru—a reprise: symbolism, function, and difference in a tibetan adept's interpretive community. Asian Music 44(1):113–139 (cit. on p. 4)
7. Das Haus Rud. Ibach Sohn, Barmen - Köln : 1794–1894 ; ein Rückblick beim Eintritt in das zweite Jahrhundert seines Bestehens (cit. on p. 4)
8. Dorje R, Ellingson T (1979) Explanation of the Secret Gcod Da ma ru an exploration of musical instrument symbolism. Asian Music 10(2):63–91 (cit. on pp. 2, 3)
9. Eberhorn M et al (2017) Web based visualization software for big data X-CT volumes with optimized Data handling and workflow. In: Vandervellen P (ed) Preservation of wooden musical instruments ethics, practice and assessment. Proceedings of the 4th annual conference COST FP1302 Wood MusICK, Brüssel, 2017, pp. 149–152 (cit. on p. 12)
10. Fforde C, Hubert J (2006) Indigenous human remains and changing museum ideology. In: Layton R (ed) A future for archaeology. UCL Press, London, pp 83–96 (cit. on p. 6)
11. Guidiance for the Care of Human Remains in Museums, London, 2005 (cit. on p. 6)
12. Henessy K (2009) Virtual repatriation and digital cultural heritage. Anthropol News 50(4):5–6 (cit. on p. 8)
13. ICOM code of ethics for museums (2017), Paris (cit. on pp. 6, 7)
14. Kinsky G (1913) Kleiner Katalog der Sammlung alter Musikinstrumente. Cöln (cit. on p. 5)
15. Kirsch S et al (2017) Some remarks on chances and challenges of computed tomography of musical instruments: the MUSICES project". In: CIMCIM bulletin 1, pp. 13–19 (cit. on p. 9)

16. Looted Art? The Benin Bronzes (cit. on p. 7)
17. Musical instruments museums online (cit. on p. 8)
18. MUSICES—Musical instruments computed tomography examination standard. Nürnberg (cit. on p. 9)
19. New virtual reality tour of the museum with oculus (cit. on p. 8)
20. Recommendations for the care of human remains in museums and collections. In: (2013) (cit. on p. 6)
21. Return of cultural objects: The Athens conference: Vol LXI, no 1–2/241–242, May 2009 (cit. on p. 7)
22. Rockhill WW (1895) Notes on the ethnology of Tibet, Washington (cit. on p. 5)
23. Sachs C (1915) Die Musikinstrumente Indiens und Indonesiens: zugleich eine Einführung in die Instrumentenkunde, Berlin (cit. on p. 5)
24. Simpson M (2009) Museums and restorative justice: heritage, repatriation and cultural education. In: Museum international LXI.1-2 (2009), pp 121–129 (cit. on p. 7)
25. The Virtual Museum of Canada: the largest digital source of stories and experiences shared by Canada's museums and heritage organizations (cit. on p. 8)
26. Wagner R et al (2018) Dual-energy computed tomography of historical musical instruments made of multiple materials. In: Proceedings of the 8th conference on industrial computed tomography, Wels, Austria (iCT 2018) (cit. on p. 9)

3D Imaging of Musical Instruments: Methods and Applications

Niko Plath

Abstract This treatise is intended as an introduction to three-dimensional (3D) imaging for stakeholders working with musical instruments, e.g. ethnologists, musicologists, curators, and instrument builders. The work should help to find the appropriate method for a specific purpose and further highlight several possible applications for the obtained data. Firstly, three techniques of 3D image acquisition are introduced. Advantages and disadvantages of the proposed methods are discussed in terms of ease of use and obtained information. Secondly, a workflow is presented to post-process the captured raw data. Finally, several examples of possible utilization of the generated virtual models are introduced. As far as possible, the proposed procedures are based on the use of open source software/freeware and should be applicable on regular current personal computers. Parts of this work are based on a proceedings abstract published in 2017 [29].

1 Introduction

3D imaging is used today for analysing, preserving and visualizing cultural heritage (CH) assets and has constantly gained in importance over the past twenty years (see Fig. 1) [34]. The prevalence of these methods is directly related to the technical development and advancements in digital electronics [18, 26, 28].

Three dimensional documentation of objects of cultural heritage can be understood as part of a natural progression from drawings, to two-dimensional photography, to 3D digital models [13]. Today, 3D images of musical instruments are generated to plan restorations, visualize functionality in education or to estimate age and origin of wooden parts. The following chapters are intended as an introduction into methods and applications for researchers, curators and instrument builders.

Described workflows are intended to run on current personal computers, only a few hard-ware requirements have to be considered: due to the size of the generated datasets, working memory is most crucial; at least 16 GB of memory is needed,

N. Plath (✉)
Institute for Systematic Musicology, Neue Rabenstr. 13, 20354 Hamburg, Germany
e-mail: niko.plath@uni-hamburg.de

R. Bader (ed.), *Computational Phonogram Archiving*, Current Research in Systematic Musicology 5, https://doi.org/10.1007/978-3-030-02695-0_15

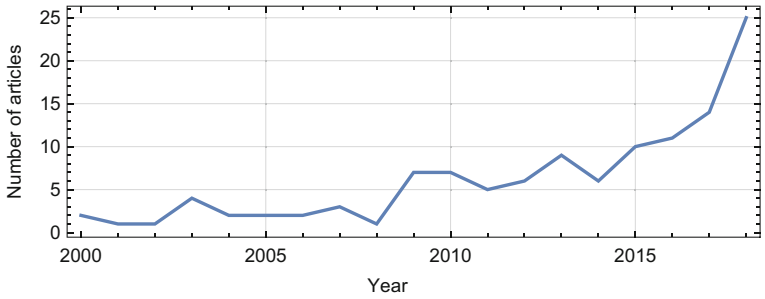


Fig. 1 Number of research articles containing "3D" in the title, per year for the *Journal of Cultural Heritage* (ISSN: 1296-2074)

32 GB is preferred. Since several processes challenge the graphics card, a standalone graphics processing unit (GPU) leads to smoother image navigation and decreased rendering time. All given information is considering the technical disposability in 2018. Suggested programs are (as far as possible) open source/freeware and work on Windows and Linux. All quoted internet addresses for software have been accessed in June 2018.

2 Methods

2.1 Photogrammetry

The most simple way to capture a 3D representation of a complex geometry like a musical instrument is to use a photogrammetric imaging technique like *Structure from Motion (SfM)* [21]. These sets of algorithms generate three-dimensional structures from sequences of two-dimensional images. Photogrammetry has been used in the field of archaeology since the nineties to capture heritage sites and as of today is also being used on a large scale on smaller objects in Museums [33]. Capturing of images can be done with a single-lens reflex camera or at the minimum with a smart phone camera in video mode [4]. Note that the spatial resolution of the generated model is directly dependent on the pixel resolution of the camera sensor. Since for the acquisition of raw image data only a simple camera is needed, it is the most mobile and least invasive (physically and socially) of the presented methods. Taking the sufficient amount of pictures (approx. 100–250) shouldn't take more than 20–30 min. Note that even if photogrammetry leads to textural object information, it is in fact capturing the apparent reflected colour without considering illumination and reflectance properties of the surface. To avoid visible signs of the lighting source on the reconstructed model (shadows, overexposed spots) indirect lighting should be used. Standards for colour acquisition in 3D imaging are yet to be defined. Since the reconstructed model contains no realistic scale, a reference measure like a folding

meter stick (or anything at hand with a known length) should be incorporated into the measuring volume. Open source solutions for the 3D mapping process are available, e.g. VisualSFM [1] or SF3M.[2] In an easy-to-use sequence of process steps, VisualSFM accomplishes matching of the images, generation of a sparse point cloud, transfer to a dense point cloud, mesh- and texture map generation.

2.2 3D Surface Scanning

Several acquisition methods are bundled under the term "3D surface scanning". Laser scanners capture surfaces by taking numerous distance measurements utilizing the time-of-flight principle. Technologies based on stereo vision reconstruct a 3D shape comparable to human vision: by calculating the distance of each captured pixel to two cameras with known position and properties. With structured light scanning, stripe patterns are projected on an object, pattern deformations are recorded by one or several cameras. Monochromatic light is usually utilized for the projection, therefore no texture information is obtained by this method. The technique allows acquisition with high geometric accuracy (20 µm or less) and measures actual sizes (direct approach to real scale). If not prohibited by conservational regulations, attaching tracking points can simplify the registration process, a subsequent merge of multiple scan series. High resolution scanners are usually mounted on a heavy camera crane, lighter handheld devices trade mobility for a lower spatial resolution.

Like any 2D camera, light based scanners can only detect surfaces which are not obscured. Therefore, instruments without air cavities like bells, gongs or cymbals are good examples for an appropriate usage of a surface scan (see Fig. 4). For hollow instruments with resonating air cavities like most string- and wind instruments a scan would only capture the outer surface.

Optical methods like photogrammetry or structured light scanning have difficulties capturing surfaces which are transparent, mirrored or polished, as well as for materials with indeterminate surfaces made from feathers, hair or felt. For industrial applications on mass-produced objects it is possible to apply removable anti-glare sprays to produce monochromatic opaque coatings. However, since this treatment should be rarely in accordance to preservation standards, it is not a convenient technique in the CH field. For scanning technologies which focus only on the object shape and ignore its optical appearance, high-dynamic-range imaging (HDRI) methods can be utilized to reduce blind spots due to shiny materials.

Figure 2 (right) shows a high resolution model of a Kulepa Ganeg, also referred to as Launut or Nunut, obtained by a structured light surface scan. The friction idiophone from Papua New Guinea is carved out of a single log and is ornamented to represent an animal (e.g. a pangolin, anteater or a bird). The three tongues are rubbed by hand producing short squeaky tones (see Fig. 2 left). The instrument is

[1] http://ccwu.me/vsfm/.

[2] http://sf3mapp.csic.es/.

Fig. 2 Kulepa Ganeg, a friction idiophone from Papua New Guinea. Depiction of a launut player on a stamp from 1979 (left). High resolution surface model based on a 3D scan (right). The original instrument is on display at Übersee-Museum Bremen, Germany

played at memorial ceremonies [36]. By means of structured light scanning, a high resolution surface model of the instrument can be created. By weighing, the average wood density could then be determined due to the exact measured object volume. For physical modelling the wood anisotropy could be approximated assuming rotational symmetry of the tree trunk.

2.3 X-Ray Computed Tomography

X-ray computed tomography (CT) has become established as a non-destructive imaging method for the examination of cultural objects [32, 41]. Biggest advantage compared to the previously presented methods is the possibility to capture measures of the complete body and thereby allowing investigation of otherwise not accessible inner parts of the object.

Medical X-ray CT facilities have been utilized to examine musical instruments since the nineties [8, 38]; lately the MUSICES project worked on establishing standard procedures for the implementation of industrial X-ray CT measurements on musical instruments [24]. High resolution industrial CT can achieve X-ray data with a spatial resolution of around 50 µm which opens the method for many fields of research, but also presents challenges regarding post-processing. One of these challenges is the required high-performance computational resources to process the achieved amount of data: Scanning of an entire violin with a spatial resolution of 100 µm yields a reconstruction consisting of up to 8000 single images with 2.5 megapixels resolution which can have an overall cumulative size of more than 60 GB.

Computed tomography based on neutron radiation can be a suitable method for materials like lead, which are not sufficiently penetrable by X-ray radiation [27]. Latest technical development include the acquisition of spatiotemporal models (4D CT), utilized in medical imaging for body parts which move throughout the measurement due to respiration.

3 Post-processing

Dependent on the method of acquisition, raw data may be present in different forms after measurements. With photogrammetry and surface scanning, usually point clouds in a 3D coordinate system are generated. Using surface reconstruction algorithms these point clouds are converted into polygonal meshes [6]. The process of computed tomography leads to a set of 2D graphics representing X-ray attenuation coefficients via grey values. These cross sectional slices may be stacked and then display a 3D volume. A common file format for storage of CT data is DICOM (Digital Imaging and Communications in Medicine), which is a non-proprietary imaging standard. The file format is widely accepted, and compared to proprietary formats a lot of free viewing and processing software is available.

To minimize the required computational resources, the cross-sectional images can be rotated and cropped in a way that they show only the proposed region of interest (ROI). This can be batch-processed with script languages like Python or with image editors like IrfanView[3] which supports handling of DICOM files via a plugin. To preserve dimensional units, note to carry the pixel spacing information from the imported header when exporting the edited cross sections. For a first approach, the images may also be down sampled by a factor of 2 (reduces the required memory to $1/2^3$).

Segmentation allows to distinguish and extract sub-structures with homogeneous properties from a heterogeneous object. It is widely used in medical applications to discriminate tissues or to generate content-related regions in machine vision. Open-source medical image analysis software is available to a great extent and can be utilized for the examination of musical instruments. There are comprehensive and powerful programs like 3DSlicer[4] or ImageJ[5] available but also small applications like ITK-Snap[6] which is specialised on the segmentation process. All mentioned programs provide semi-automatic segmentation processing (based on region growing) and perform well even with large-scale data sets [43]. Each voxel is allocated to exactly one of the defined sub-volumes, therefore different sub-volumes cannot overlap, which is crucial for subsequent applications like physical modelling or rapid prototyping. Note that filtering or re-meshing the individual sub-volumes can destroy this ratio. Since sub-volumes are generated according to their homogeneous properties, the respective surface data adequately describes a body (the question which degree of heterogeneity is interpreted as to be homogeneous depends on the research question). This leads to a massive reduction of data. Handling, therefore, is simplified: 3D meshes of musical instrument parts can easily be sent per email and individual sections can be stored in databases for specific research tasks.

For the exact determination of the surface, a scan with as less noise and other image errors (e.g. metal artefacts) as possible is necessary. However, even for an

[3] https://www.irfanview.com/.

[4] https://www.slicer.org/.

[5] https://imagej.net/.

[6] http://www.itksnap.org/.

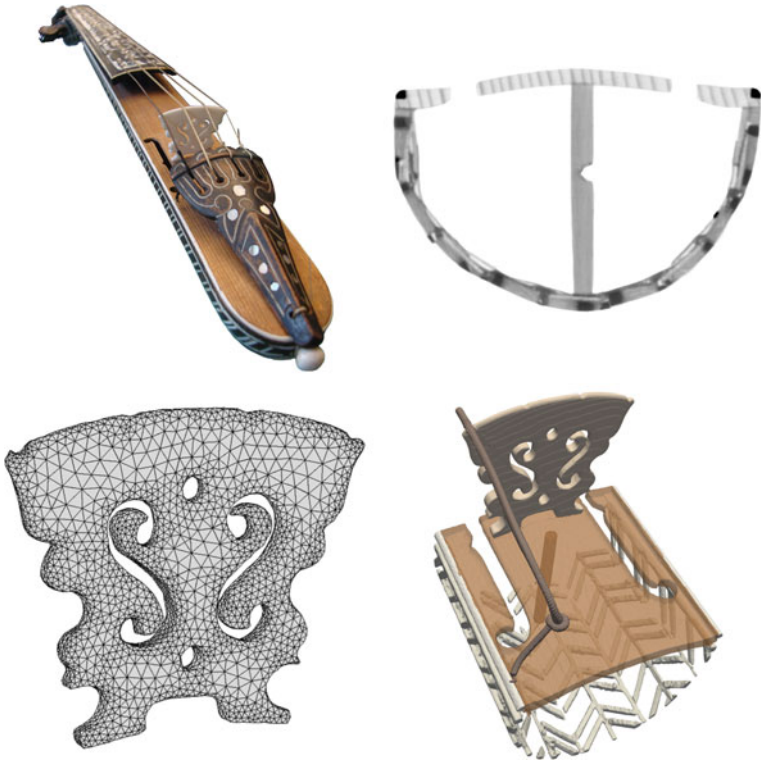


Fig. 3 Exemplary results of the proposed workflow (clockwise): photo of a 17th century kit violin, cross-sectional CT image (inverted grey values), rendering of sub-structures after segmentation, finite element model generated from the bridge sub-volume. The original instrument is on display at Germanisches Nationalmuseum Nürnberg, Germany. The CT data has been acquired by the MUSICES project [24]

excellent scan the exported surface meshes (most commonly stored as .stl or .obj files) may have diverse defects, e.g. self-intersecting faces, non-manifold edges or unreferenced vertices. The software MeshLab[7] is recommended for this purpose, as it provides various automated mesh cleaning filters. A reliable way to obtain a clean and watertight mesh is a Poisson surface reconstruction (PSR). But note that since this is actually a re-meshing, individual sub-volumes could overlap afterwards [22]. In general, editing steps like smoothing, down sampling or mesh simplification could substantially modify or even destroy data and therefore must be documented in a transparent and reproducible way.

Figure 3 shows intermediate results of the proposed workflow. A kit violin is scanned with high resolution X-ray CT. Segmentation of the cross sectional slices leads to sub-volumes of all constructional instrument parts. Subsequently, individual parts can be reproduced, measured, or utilized for physical modelling.

[7] http://www.meshlab.net/.

After segmentation and classification of sub-volumes, the geometries are usually stored as surface meshes and information about the inner structure of each sub-volume is lost. For wooden instrument parts, it is desirable to preserve some information about the inner structure, e.g. wood density and grain directions. A simple workaround to retain the directivity can be to manually add a second solid into the respective .stl file. This solid only contains vertices forming three orthogonal unit vectors from the origin for the wood directions *longitudinal* (l), *radial* (r) and tangential (t):

```
solid directions
    facet normal ni nj nk
        outer loop
            vertex v_lx v_ly v_lz
            vertex v_rx v_ry v_rz
            vertex v_tx v_ty v_tz
        endloop
    endfacet
endsolid directions
```

4 Applications

4.1 Geometric Measures

Geometric measures can be made directly on the 3D model or can be derived from containing information. The results can provide useful insight for further research or technical drawings. The thickness distributions of soundboards, dimensions of restored patches or the volume of complex geometries like a bell can be measured directly from the model. By using wooden calibration bodies, absolute wood densities can be measured (see Sect. 4.2). In combination with "external" information, e.g. the weight, wood species or metal alloy, more information can be derived: Since it is known that the violin bridge in Fig. 3 is made from ivory, measuring the body volume (896 mm^3) from the virtual model and estimating its density based on literature resources (1700–1900 kg/m^3) the bridge weight can be derived without disassembly (1.5–1.7 g).

4.2 Dendrochronological Dating and Density Measurements

Cross sections of CT data can be utilized for dendrochronological dating [35]. The required resolution depends on the smallest distance between annual rings, but a spatial resolution of 80 μm is usually sufficient. Cross sections as a basis are ideally suited when certain wooden parts are covered with dark varnish or the part to be

Table 1 Species selection for a wood density reference object when performing X-ray CT based density measurements of violin family instrument parts. Average densities from [9]

Instrument part	Species	Density (kg/m^3)
Top plate	Picea abies	445
	Picea sitchensis	424
	Picea engelmannii	385
	Thuja plicata	370
Back plate, ribs, neck	Acer platanoides	750
	Acer saccharinum	760
	Acer pseudoplatanus	630
	Acer pseudoplatanus	580
Bow	Paubrasilia echinata	980
	Brosimum guianense	1210
Bridge	Acer campestre	690
	Acer saccharum	705
Fingerboards, pegs, tailpieces	Diospyros crassiflora	955
	Buxus sempervirens	975
	Dalbergia nigra	835
	Cercocarpus spp.	1110

analysed is implemented in the instrument like in the case of a patch under the soundboard of a violin.

The obtained attenuation coefficients—transferred to grey values of voxels in a CT cross sectional image—are dependent on the energy used for the X-ray exposure and on the density distribution in the object under investigation. Therefore, they describe only relative ratios of density which have to be scaled to absolute values by means of reference regions with known density, e.g. air or water in medical applications [40]. In the case of wooden musical instruments, knowledge about absolute wood density is an important information for reproduction, restoration or physical modelling of an instrument. Wood shows a linear relation between absolute density and mean attenuation coefficient [17]. Therefore, it is a common approach to place several wooden reference samples with known density into the measuring volume for comparison [42]. Table 1 depicts a proposal for reference wood samples covering the density range of all wood species used in violin family instruments (approx. 440–1200 kg/m^3). Small cubic wood samples are arranged in a bar shape and placed into the measuring volume orthogonal to the X-ray beam. This ensures that the beam only crosses one species on its way to the detector and diffraction error is minimized. The gradient of the relation between attenuation coefficient and absolute density can be determined by linear extrapolation and then used to scale grey values to absolute density.

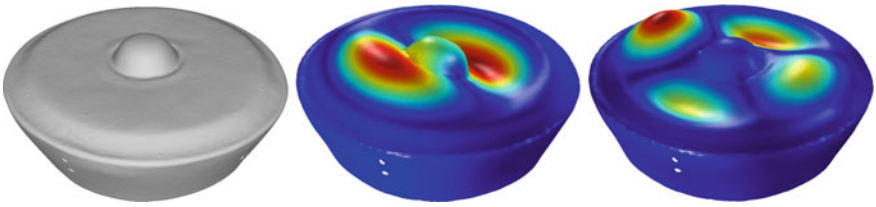


Fig. 4 One gong of a Philippine Kulintang chime ensemble. The geometric model based on a surface scan is utilized for finite element method (FEM) physical modelling. Modal deflection shapes ([0, 1] and [1, 1]) obtained from an eigenvalue analysis with free-free boundary conditions. The original instrument is situated at the University of Hamburg, Germany

4.3 Physical Modelling

Sub-volume data can be utilized for physical modelling, thus allowing the evaluation of the vibroacoustic behaviour of instruments, even in non-playable condition [31]. With regard to conservation issues, physical modelling can also provide insights into the structural behaviour of displayed instruments under long term mechanical loading [25].

Figure 4 shows a model of a Philippine Kulintang bronze gong based on a structured light scan. Assuming a homogeneous alloy, weighting and measurement of the exact volume provides the material density. As a first approach, Young's modulus and Poisson's ratio are approximated based on literature. The surface model is transformed into a solid body and re-meshed into a combination of simple geometric shapes ("finite elements").

The 3D modelling software FreeCAD[8] provides a module to perform simple finite element analysis (FEA), OpenFOAM[9] is mainly a tool for computational fluid dynamics but also includes FEA. Certainly, to get scanned complex geometries with heterogeneous material properties working well in physics simulation software is still a challenging task [12].

4.4 Rapid Prototyping

Since in many cases historical instruments are not in playable condition, the possibility to produce exact copies can be of great value. Using extracted surface data of CT scans, whole instruments can be reproduced with high geometric precision as well as missing parts for restoration purposes [37]. Manufacture may be obtained by additive (3D printing) or subtractive methods (CNC milling). Several consumer 3D printers based on fused deposit modelling (FDM) can achieve a vertical spatial resolution of 100 µm, printers based on stereolithography (SLA) realise comparable

[8] https://www.freecadweb.org.

[9] https://www.openfoam.com/.

or higher resolutions in all axes. From a vibroacoustical point of view the substitution of complex materials like wood is difficult. Currently, substitution of natural materials by composites is intensively studied and of great interest for musical instruments builders. Although it is possible by now to print composites of wood and polyester, the exact orthotropic material properties are not yet reproducible. For stakeholders using rapid prototyping methods, the most important factors are surface quality, dimensional accuracy and static tensile mechanical behaviour. As of yet, the vibroacoustic behaviour of the base material is of no importance (advertising slogans for printing material promote an "attractive wood-like look" or a "metal-like finish" as opposed to a "wood-like vibrational behaviour"). The same is also true for the composition of printed infill structures: triangle, rectangular or honeycomb shapes are generated by slicer applications to maximize tensile strength with minimum object weight. Again, no considerations are made regarding the vibrational behaviour of the object. Nevertheless, rapid prototyping can be utilized e.g. during pre-planning of restoration projects or to aid the understanding of complex mechanical details, in workshop and exhibition [2]. From a conservation perspective, studies about long-term stability and chemical usability of the prevalent polymer compositions are still pending [14].

4.5 Virtual/Mixed Reality Applications

Augmented-, virtual- and mixed reality (AR/VR/MR) applications give museums novel possibilities to impart knowledge about cultural artefacts [5, 11, 19]. Immersive technologies can be utilized in exhibitions, education, restoration and for virtual tours. The use of VR technologies can not only increase the actual number of people who have access to a certain object drastically but also improves the quality of interaction with the object: Fig. 5 depicts a view into a VR rendering of a viola. The observer can explore the inner parts of the object and otherwise hidden complex mechanical functionality can be inspected. Annotations can help to guide through the object via text, sound or video.

4.6 Digital Archiving

Today, stakeholders from various fields like musicology, cultural history, ethnology, conservation, engineering and instrument building interact with digital 3D representations of musical instruments. Archival structures, therefore, need to be able to handle heterogeneous sets of sources, data structures, content and formats and thereby allow interdisciplinary approach to the objects. The output formats of datasets vary widely: raw geometric 3D models, web based applications, renderings, animations, VR and AR applications, annotations, 3D sketches, vector graphics and technical drawings among others. As of today, there is no single solution to cover all these

Fig. 5 VR view into the rendered inner part of a viola. The observer can "walk" through the instrument, individual parts are annotated to explain their functionality. The model is based on an X-ray CT scan, the rendering is generated with the VR viewer from Sketchfab. Exemplary annotation text retrieved May 4, 2018, from https://en.wikipedia.org/wiki/sound_post

demands but a variety of services, applications and databases is used for each specific task:

In the past, Geographic information systems (GIS) have been used extensively to examine CH objects, allowing to handle, analyse and annotate 3D models. Existing CH software applications like SICaR [3], Hyper3D [23] or Neptune IS [1] are specialized on assisting restoration processes. Web-based collections of European CH objects start to embed 3D models, although, interaction with the 3D content is yet very limited.[10] The worlds largest online database of musical instruments held in collections[11] does not yet provide support for 3D models but can handle CT cross sections. Recently, technical frameworks for web-based visual presentation of high resolution 3D images are developed [15, 16, 30]. Multiresolution encoding should allow web-based interaction with large models of millions of polygons.

Further crucial challenges for archiving structures are the following:

- A standard documentation format for the entire data processing pipeline has to be established. Post-processing 3D data includes steps which can drastically alter the appearance of the final model, e.g. chosen thresholds for surface reconstruction, remeshing or threshold selections for segmentation of sub-volumes [7]. Therefore, these processes should be documented in a transparent, reproducible way.

[10] https://www.europeana.eu.

[11] http://www.mimo-international.com.

- Data exchange formats should be in compliance with existing standards (e.g. CIDOC Conceptual Reference Model).
- A framework is needed which allows to combine data from the tangible and intangible domain [10].
- It should be possible for the non-expert to interact with large polygonal models of millions of polygons.
- It should be possible for the stakeholder to contribute to the datasets (e.g. with metadata, annotations and sketches on 3D models [20, 39]).

To allow for archives to advance from being mere storage facilities towards being comprehensive research tools, interaction with 3D content should be facilitated and encouraged. 3D visual representations of historical artefacts provide an intuitive and immersive environment, encouraging novel and creative approaches to cultural heritage safeguarding.

References

1. Apollonio FI et al (2018) A 3D-centered information system for the documentation of a complex restoration intervention. J Cult Herit 29:89–99 (cit. on p. 11)
2. Balletti C, Ballarin M, Guerra F (2017) 3D printing: state of the art and future perspectives. J Cult Herit 26:172–182 (cit. on p. 10)
3. Baracchini C et al (2003) SICAR: geographic information system for the documentation of restoration analysis and intervention article. In: Salimbeni R (ed) Proceedings of SPIE—the international society for optical engineering, vol 5146, pp 149–160 (cit. on p. 11)
4. Barbero-García I et al (2018) Smartphone-based close-range photogrammetric assessment of spherical objects. Photogramm Rec 33(6):283–299 (cit. on p. 2)
5. Bekele MK et al (2018) A survey of augmented, virtual, and mixed reality for cultural heritage. J Comput Cult Herit 11(2):1–36 (cit. on p. 10)
6. Berger M et al (2014) State of the art in surface reconstruction from point clouds. In: Eurographics 2014-state of the art reports 1, pp 161–185 (cit. on p. 5)
7. Bernardini F, Rushmeier H (2002) The 3D model acquisition pipeline. Comput Graph Forum 21(2):149–172 (cit. on p. 11)
8. Borman T, Stoel B (2009) Review of the uses of computed tomography for analyzing instruments of the Violin family with a focus on the future. J Violin Soc Am XXII(1):1–12 (cit. on p. 4)
9. Bucur V (2006) Acoustics of wood. Springer (cit. on p. 8)
10. Carboni N, de Luca L (2016) Towards a conceptual foundation for documenting tangible and intangible elements of a cultural object. Digit Appl Archaeol Cult Herit 3(4):108–116 (cit. on p. 12)
11. Carrozzino M, Bergamasco M (2010) Beyond virtual museums: experiencing immersive virtual reality in real museums. J Cult Herit 11(4):452–458 (cit. on p. 10)
12. Cepeda JF et al (2013) A practical method to model complex threedimensional geometries with non-uniform material properties using image-based design and COMSOL multiphysics. In: Proceedings of the 2013 COMSOL conference in Boston (cit. on p. 9)
13. Cignoni P, Scopigno R (2008) Sampled 3D models for CH applications: a viable and enabling new medium or just a technological exercise? J Comput Cult Herit 1(1):1–23 (cit. on p. 1)
14. Coon C et al (2016) Preserving rapid prototypes: a review. Herit Sci 4(1):40 (cit. on p. 10)

15. Eberhorn M et al (2017) Web based visualization software for big data X-CT volumes with optimized datahandling and workflow. In: Preservation of wooden musical instruments—4th annual conference COST FP1302 WoodMusICK. Ethics, practice and assessment, pp 149–152 (cit. on p. 11)
16. Guarnieri A, Pirotti F, Vettore A (2010) Cultural heritage interactive 3D models on the web: an approach using open source and free software. J Cult Herit 11(3):350–353 (cit. on p. 11)
17. Heismann BJ, Leppert J, Stierstorfer K (2003) Density and atomic number measurements with spectral X-ray attenuation method. J Appl Phys 94(3):2073–2079 (cit. on p. 8)
18. Ioannides M, Quak E (eds) (2014) 3D research challenges in cultural heritage. Lecture notes in computer science, vol 8355. Springer, Berlin, Heidelberg, p 151 (cit. on p. 1)
19. Jiménez Fernández-Palacios B, Morabito D, Remondino F (2017) Access to complex reality-based 3D models using virtual reality solutions. J Cult Herit 23:40–48 (cit. on p. 10)
20. Jung T, Gross MD, Do EY-L (2002) Annotating and sketching on 3D web models. In: Proceedings of the 7th international conference on intelligent user interfaces—IUI'02. ACM Press, New York, USA, p 95 (cit. on p. 12)
21. Katz J (2017) Digitized Maya music: the creation of a 3D database of Maya musical artifacts. Digit Appl Archaeol Cult Herit 6:29–37 (cit. on p. 2)
22. Kazhdan M, Bolitho M, Hoppe H (2006) Poisson surface reconstruction. In: Proceedings of the symposium on geometry processing, pp 61–70. arXiv:1006.4903 (cit. on p. 6)
23. Kim MH et al (2014) Hyper3D. J Comput Cult Herit 7(3):1–19 (cit. on p. 11)
24. Kirsch S et al (2017) Some remarks on chances and challenges of computed tomography of musical instruments. The "MUSICES" project. CIMCIM Bull 1 (cit. on pp. 4, 6)
25. Konopka (2016) Hygro-mechanical structural analysis of keyboard instruments. In: Analysis and characterisation of wooden cultural heritage by scientific engineering methods, pp 65–71 (cit. on p. 9)
26. MacDonald L (2006) Digital heritage. Routledge, pp 448–463 (cit. on p. 1)
27. Mannes D et al (2015) Combined neutron and X-ray imaging for noninvasive investigations of cultural heritage objects. Phys Procedia 69(69):653–660 (cit. on p. 4)
28. Pears N, Liu Y, Bunting P (eds) (2012) 3D imaging, analysis and applications. Springer, London (cit. on p. 1)
29. Plath N, Kirsch S (2017) Post-processing of musical instrument 3D-computed tomography data for conservational applications. In: Preservation of wooden musical instruments—4th annual conference COST FP1302 WoodMusICK, pp 161–164 (cit. on p. 1)
30. Potenziani M et al (2015) 3DHOP: 3D heritage online presenter. Comput Graph 52:129–141 (cit. on p. 11)
31. Pyrkosz M, Van Karsen C, Bissinger G (2011) Converting CT scans of a Stradivari Violin to a FEM. In: Proulx T (ed) Proceedings of the IMAC-XXVIII. Conference proceedings of the society for experimental mechanics series, vol 3. Springer, New York, pp 811–820 (cit. on p. 9)
32. Re A et al (2014) X-ray tomography of large wooden artworks: the case study of "Doppio corpo" by Pietro Piffetti. Herit Sci 2(1):19 (cit. on p. 4)
33. Remondino F (2011) Heritage recording and 3D modeling with photogrammetry and 3D scanning. Remote Sens 3(6):1104–1138 (cit. on p. 2)
34. Remondino F et al (2009) 3D modeling of complex and detailed cultural heritage using multi-resolution data. J Comput Cult Herit 2(1):1–20 (cit. on p. 1)
35. Rigon L et al (2010) Synchrotron-radiation microtomography for the non-destructive structural evaluation of bowed stringed instruments. E-Preserv Sci 7:71–77 (cit. on p. 7)
36. Sachs C (1913) Real-Lexikon der Musikinstrumente, zugleich Polyglossar für das gesamte Instrumentengebiet. Berlin, Julius Bard (cit. on p. 4)
37. Savan J, Simian R (2014) CAD modelling and 3D printing for musical instrument research: the Renaissance cornett as a case study. Early Music 42(4):537–544 (cit. on p. 9)
38. Sirr SA, Waddle JR (1997) CT analysis of bowed stringed instruments. Radiology 203(3):801–805 (cit. on p. 4)

39. Soler F, Melero FJ, Luzón MV (2017) A complete 3D information system for cultural heritage documentation. J Cult Herit 23:49–57 (cit. on p. 12)
40. Stoel BC et al (2012) Wood densitometry in 17th and 18th century Dutch, German, Austrian and French Violins, compared to classical Cremonese and modern Violins. PLoS ONE 7(10) (cit. on p. 8)
41. Van den Bulcke J et al (2017) Nondestructive research on wooden musical instruments: from macro- to microscale imaging with lab-based X-ray CT systems. J Cult Herit 27:S78–S87 (cit. on p. 4)
42. Wei Q, Leblon B, La Rocque A (2011) On the use of X-ray computed tomography for determining wood properties: a review. Can J For Res 41(11):2120–2140 (cit. on p. 8)
43. Yushkevich PA et al (2006) User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. NeuroImage 31(3):1116–1128 (cit. on p. 5)

How to Interprete Early Recordings? Artefacts and Resonances in Recording and Reproduction of Singing Voices

Malte Kob and Tobias A. Weege

Abstract Voice recordings in the beginning of the 20th century required a complex acoustic and mechanical set-up for transfer of the singing voice sound from singer via horns, ducts, soundbox and needle to a cylinder or disc. A similar construction was used to reproduce the voice. Both signal paths had significant impact on various properties of the voice signals—mostly the voice signal quality was reduced and distortions were produced. Another effect were changes in details of the voice spectrum due to the interaction of resonances in the voice signal and the transfer paths of recording and reproduction devices. We present analyses of the voice signal modifications and relate them to parts of the devices. This work is funded by the German Research Foundation.

1 Introduction

The invention of sound recording technology, in the second half of the 19th century, opened a completely new horizon for studying the musical practice of different cultures. Especially the phonograph and the associated wax cylinder technology, that constituted a quite robust and easy to handle field recording setup, quickly became one of the most important tools and sources for ethnomusicological studies.

The many elements that lie between the original sound that was produced in front of the recording horns, sometimes more than a century ago, and the sound that reaches the ears of a contemporary researcher can substantially influence the information retrieval process. The question of conservation of the original sound carrier and digitization with the best fitted technique is complex enough already. But even when using the best available techniques for these processes, there are other

M. Kob (✉) · T. A. Weege
Detmold University of Music, 32756 Detmold, Germany
e-mail: kob@hfm-detmold.de

R. Bader (ed.), *Computational Phonogram Archiving*, Current Research in Systematic Musicology 5, https://doi.org/10.1007/978-3-030-02695-0_16

aspects that can influence the conclusions drawn from listening to an antique audio track.

An attentive listener will immediately notice the limited frequency-band of an old recording, compared to a modern one. What not many listeners are aware of, however, is that, apart from this limited frequency-band, there can be considerable differences in the sub-bands. The resonances of the mechanical elements in the recording and reproducing chains will enhance some frequencies, while others that lie apart from resonances can be considerably suppressed. In this way, the listener will be given a colored sound with a non-flat frequency response characteristic.

Every researcher that analyses historical recordings should be aware of these phenomena, in order to avoid describing the sound source with an attribute that might have been introduced by the recording and reproducing devices. These questions are relevant not only for ethnomusicologists, but also for other researchers that are interested in the historical musical practice in western classical music, for instance. In this context, the shellac discs and the related gramophone recording and reproducing technologies play the main role.

The recording process of the gramophone technology was not really compatible with field recording situations. The need for a series of chemical processes before becoming a playable disc, for instance, limited the use by independent researchers or non-profit institutions. It was very well suited and successful, on the other hand, for commercial oriented recordings, in this way widely documenting the achievements of great musicians over the last century.

The previously mentioned sound coloring effects should also be taken in account when describing, for instance, the singing technique of an artist recorded on shellac. A ratter soft voice can be perceived as being quite heroic, depending on the frequencies that are enhanced. A ratter soft vibrato can sound quite penetrating when hitting the edge of a sharp resonance of the mechanical recording setup. In addition to the resonances of the mechanical systems, there are also the resonances of the vocal tract of a singer, known as formants, that will be superimposed to the former ones. This and other properties of the voice that can interact with the devices are the topic of Sects. 2 and 6 in this article.

Achieving a precise interpretation of early recordings shows itself as a rather complex task. A way to confront this complexity is dividing the recording and reproducing chains in smaller parts and analyzing the influence of each part on sound. This is the approach followed by the authors of this article, where the results of vibro-acoustical measurements performed on the individual parts of a gramophone and their combinations are presented and discussed. Section 3 focuses on the recording path, while the observations on the reproducing chain are presented in Sect. 4.

2 Voice Properties

Voice signals exhibit a rich set of features due to their complex generation mechanism and large variability. In the early 20th century, modern singing techniques such as belting, beatbox, or overtone singing [11], used in Pop and Jazz music today

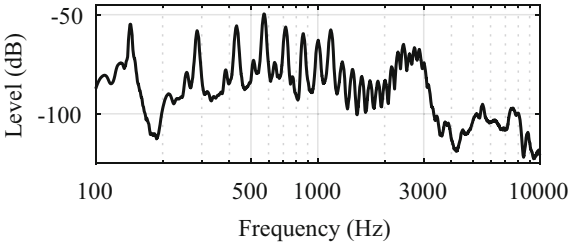


Fig. 1 Smoothed spectrum (1/24 oct.) of the vowel /a:/ with singer's formant

had not yet been invented, but singers used various techniques to increase expressiveness and voice projection. The use of falsetto was a common method to draw attention to listeners as well as the use of vibrato and glides [5]. In operatic singing the singer's formant was used to enhance the projection of voice over large orchestras [17]. Vibrato and glides were ornaments that not only musically enriched the performance [4] but also helped to increase the audibility of recorded voices in the presence of artefacts such as noise, crackle and distortion which have a masking effect especially on soft voices.

The voice signal contains frequency components from the low fundamental below 80 Hz up to frequencies above the singer's formant around 4 kHz up to the upper limit of human hearing. The voice timbre is significantly determined by the choice of the vowel in voiced phonemes (see Fig. 1) whereas consonants feature transient and non-harmonic signal components in a broad frequency range.

The distribution of the signal components vary strongly with the singing style and the fundamental frequency. Since the variation of the fundamental frequency immediately changes all harmonics of the voice sound accordingly, vibrato, ornaments and glides will affect the whole spectrum of voice signals. This has important consequences for the interaction of the voice signal and the properties of recording or reproduction devices as discussed in the next sections.

3 Transfer Path from Singer to Disc

As mentioned in Sect. 1, the phenomena discussed in this article rely on measurements performed on gramophone recording and reproducing setups. Nevertheless, the main elements in the path of the sound during the acoustic era, independently if looking at phonograph or gramophone technology are roughly the same. In the recording path, there is a recording horn, a recording soundbox, a tube (linking the former elements), a media on which to be written, and a mechanical system providing the rotation of the media. Ideally, this last element, better known as recording lathe, would provide a constant rotational speed, thus having no influence on sound. The media on which the sound was to be written (a wax disc, for gramophone technology), on the other hand, certainly had an influence as a form of termination impedance,

Fig. 2 Replica of recording horn 11 1/2

damping the movement of the cutting stylus. This influence could not be assessed so far, and for the present work it was neglected.

The awareness of the influence of mechanical elements on sound is nothing really new. As early as 1890, an article in the Scientific American journal (as cited in [19, p. 62–63]) mentions the influence of soundbox diaphragm's modes on sound as a quite obvious and well known factor.

The influence of elements in the recording and playing chains on sound has been discussed in sparse works over the last decades [2, 3, 12, 13, 16, 20], but without reaching or even aiming for quantitatively describing the entire vibro-acoustical path. Recently some results in this direction have been presented [18] and are here discussed as well.

3.1 Influence of the Horn

The horn is the first element in the recording chain, closest to the sound source. Most recording horns had a conical form and were made out of metallic alloys. Each recording company had a set of horns with different sizes and forms that were selected for each session according to aspects like music style and instrumentation. The Gramophone Co., one of the leading companies in the acoustical era, had a set of at least 12 recording horns [2, p. 15].

The influence of a recording horn on sound can be understood on the basis of the frequency response function (FRF) shown in Fig. 3. This data is the outcome of transfer function measurements, with sweeps as input signal,[1] performed on a replica of one of the 12 aforementioned recording horns, the so called 11 1/2 horn, shown in Fig. 2.

In Fig. 3, the peaks in the frequency response correspond to the axial resonances of the conical horn. The first and strongest resonance lies around 200 Hz. If this horn was the only element between a sound source and a listener, i.e. if a listener would position the smaller end of the horn at one of his ears, he would hear a strongly colored sound. All frequency components close to 200, 400, 600 Hz, ... would be boosted, while the ones around 250, 500 Hz, ... would be strongly attenuated. Changes in the length or in the form of the horn would cause the position and the sharpness of the peaks to be changed. The utilization of different horns, and sometimes of many horns, was one of the ways, even if limited, for the recording engineer to achieve a desired

[1] For details on the measurement setups, please refer to [18].

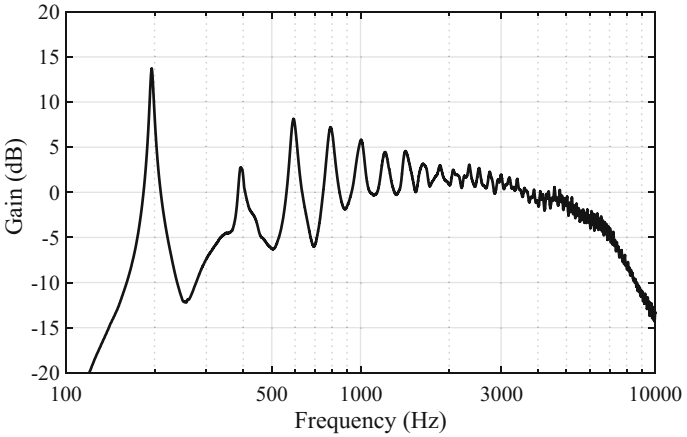


Fig. 3 FRF of the recording horn replica

equalization of the sound. When recording a solo piece played on a grand piano, with a frequency range of interest that usually goes quite lower than for voices, a longer horn would be preferred. The placement of the first axial resonance of the horn at a lower frequency would provide some boost in this range.

3.2 Influence of the Tube

All recording setups need an element linking the horn to the recording soundbox.[2] Not much is known about this part, mainly that they were made out of rubber and were commonly about 10 cm long [3, p. 264]. Nevertheless, as already pointed out by Copeland [3, pp. 265–266], their influence should not be underestimated. This influence will be discussed on the basis of the frequency response functions obtained for tubes of different lengths connected to the recording horn.

Starting with the FRF of the horn only (as already shown in Fig. 3) on the very top, Fig. 4 comparatively shows the results obtained when connecting soft PVC tubes of three different lengths to the horn's throat. Attaching the 5 cm tube barely affects the FRF, in comparison to the response for horn only. The central frequency of the resonances is only slightly lowered and a soft modulation is added starting from about 500 Hz. When attaching the 10 cm tube, apart from the fact that the central frequencies of the resonances are again lowered, the second resonance (between 300 and 400 Hz), that was quite weak in the previous setups, is considerably increased. For the 15 cm tube, this resonance is about 20 dB stronger than in the top graphic.

[2] When using more than one horn, a junction connecting the acoustic signals coming from the different horns would additionally be needed. This special cases will not be treated here.

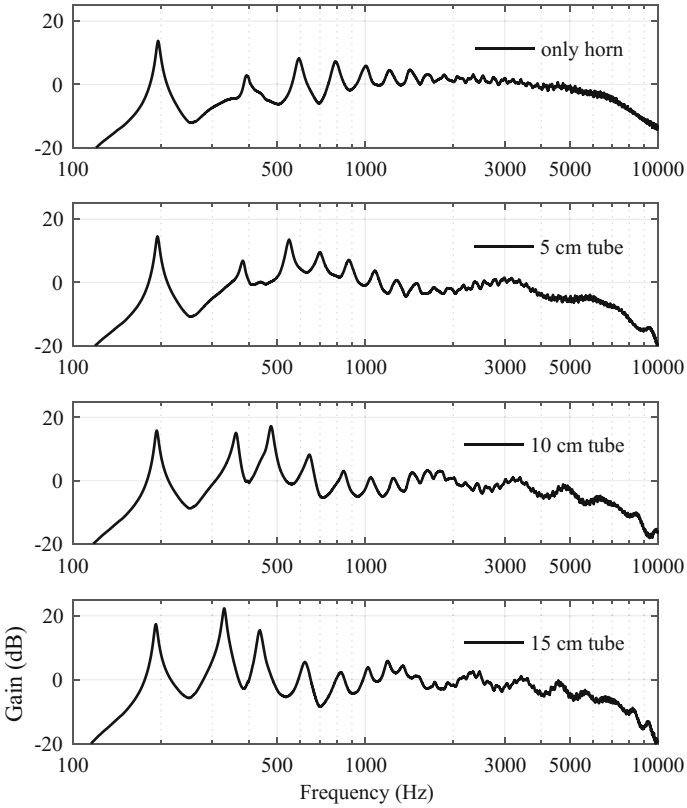


Fig. 4 FRFs of tubes with different lengths attached to the recording horn's throat

The influence evidenced in the graphs of Fig. 4 is more dramatic in a frequency range between 300 and 600 Hz. This range concentrates a lot of musical information, as the fundamental frequencies and first harmonics in the registers of many musical instruments. Also for vocal music, the formants corresponding to many vowels are concentrated in this range.

3.3 Influence of the Soundbox

The soundbox is the element responsible for the conversion from airborne to structure-borne sound. The operating principle of a recording soundbox is very similar to the more common reproducing soundboxes. The most relevant difference between reproducing and recording soundboxes seems to be in terms of the stylus-lever mechanism connecting the membrane to the cutting stylus: the stiffness at the pivot was much lower in the recording ones [3, p. 269]. As mentioned already for

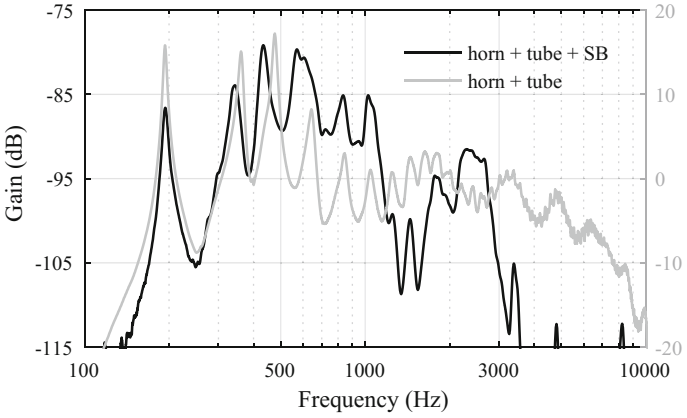


Fig. 5 FRFs of the recording chain

recording horns, it was common that the recording engineers had a set of recording soundboxes at their disposal. The variation was essentially in terms of the membrane's diameter and thickness.[3] Empirically, they selected the soundbox thought to be most adequate for each recording situation.

The recording soundboxes were a very sensitive element in the context of protecting industrial secrets and of competition between the recording companies. Only few recording soundboxes survived and it was not possible for the authors to have access to either an original exemplar or a replica so far. The influence of this element is discussed on the basis of experiments where a reproducing soundbox[4] was connected to the elements already described in Sects. 3.1 and 3.2.[5]

Figure 5 shows the FRF obtained from the measurement of the transfer function between the sound arriving at the horn and the resulting velocity of the soundbox needle. The grey line shows, for comparison, the FRF of the previous setup, with horn and tube only. In general terms, it is possible to affirm that most of the pronounced resonances that could be observed for the setup without the soundbox (grey line) are still present when adding the soundbox (black line). Although they may have been slightly softened and their central frequency may have suffered a small shift, there are no major changes observable in the range until 500 Hz in Fig. 5.

Around 600, 1000 and 2400 Hz in Fig. 5 it is possible to observe an enhancement of the resonances, while the resonances originally around 1500 Hz are strongly

[3] Typical ranges were 30–50 mm and 0.17–0.25 mm, respectively [3, p. 269].

[4] This decision, aiming for having a reasonably similar ending impedance in the recording chain, is based on the similarity of the functioning principal between reproducing and recording soundboxes. Nevertheless, the differences between them, as mentioned in the previous paragraph, would certainly influence the frequency response function in some extent, what should be investigated in further works, when guaranteed the access to an original exemplar.

[5] A 10 cm tube was used as element linking the horn replica to the soundbox, since this is reported to have been a commonly used length for this element [3], as already mentioned in Sect. 3.2.

damped. This behavior is caused by the vibrational modes of the soundbox membrane. Frequencies around the resonances of the membrane are enhanced, while the ones between resonances are damped. Another behavior introduced by the presence of the soundbox is the sharp high-cut at 3000 Hz. These phenomena will be further discussed in Sect. 4.1.

4 Transfer Path from Disc to Listener

Although essential for understanding the influence of early recording setups on sound, the discussion presented in Sect. 3 has only limited practical value if not complemented with investigations on the reproducing chain. The reason for this being quite obvious: the information contained in shellac discs or wax cylinders has to be retrieved using some method, that will always have an influence[6] on the resulting sound.

In the case of transfers from discs or cylinders, elements like the amplifier, the cartridge and the needle tip will influence the sound. The discussion of these methods is beyond the scope of this text, but it can be affirmed that an experienced audio engineer can minimize and/or compensate the influence of his reproducing setup, resulting in a situation where the "fingerprint" of the recording chain will dominate the sound coloring. Sometimes the engineer will try to compensate for this influence of the recording setup.

There is an aspect, however, that will influence the sound in the case of transfers as well as in mechano-acoustical setups: the reproduction rotational speed. Researchers that intend to draw conclusions about musical practice while listening to historical recordings should be aware of this aspect, that influences the sound and can bias their opinions. Although already being discussed in the 1980s [1], it is still a relevant topic.

The main goal of this section is to investigate the influence of a mechano-acoustical reproducing setup on sound. In other words, to investigate how the devices originally used for playing the wax cylinders and shellac discs influenced the sound. Nowadays, most collectors of shellac discs do not even want to incur the risk of wearing their treasures by playing them back on a gramophone. For many decades of the last century, however, these devices and all their secondary effects on sound dominated the recorded music soundscape.

The main elements of a gramophone that are acoustically relevant are the soundbox, the tonearm and the horn.[7] The spring motor of a gramophone was an important factor, dividing low and high performance devices. In the ideal case, however,

[6] Optical digitizing methods, that were developed already in the 1980s [6, 7], and could considerably minimize this influence still have many limitations, which discussion is beyond the scope of this text.

[7] The influence of the needle type on sound is perceptible and relevant, but introduces an enormous variance factor that was not part of this study, where only a medium tone steel needle was used.

Fig. 6 Academy gramophone

analogously to what was already discussed for the recording lathe, it would provide a constant rotation during an entire playing period, thus having no influence on sound.

As was the case for the recording chain, in Sect. 3, the here presented measurement results will also be divided in subsections, each one corresponding to a part that is added to the setup. The measured parts are from a gramophone of the English manufacturer Academy, shown in Fig. 6.

4.1 Influence of the Soundbox

Last element in the path from sound source to media in the recording chain, the soundbox is the first element in the reproducing chain. Inversely to the conversion that takes place in a recording soundbox, here it is the element responsible for converting the structure-borne sound, produced through the displacement of the needle tip according to the grooves of the disc or cylinder, to airborne sound that will be fed to the tonearm, horn, room, and listener.

As was the case when investigating the recording chain (Sect. 3), the frequency response functions obtained from transfer function measurements with sweeps as input signal will build the basis for the discussion. Here, the input to the system is the movement of the needle tip.[8]

The soundbox part where the conversion from structure-borne to airborne sound actually occurs is the membrane or diaphragm. The vibrational modes of the

[8]This movement is emulated with a shaker. The input velocity is tracked by means of an accelerometer. For details on the measurement setups, please refer to [18]. There can also be found the results of comparative measurements on 4 models of soundboxes.

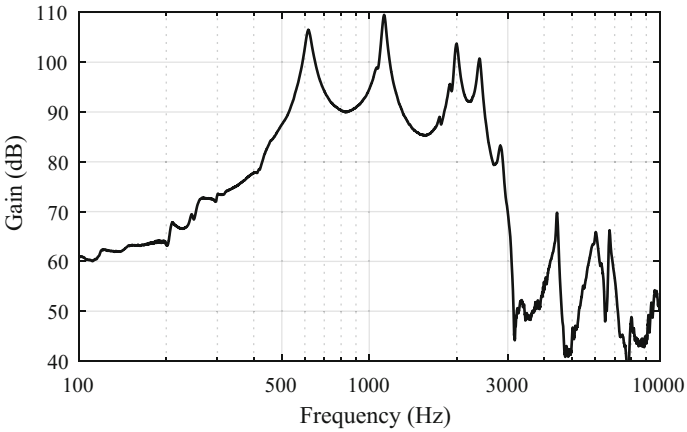


Fig. 7 FRF of the soundbox

membrane have an influence on sound [3, 13, 16, 20], enhancing the frequencies around the resonances. This "fingerprint" of the modes was already observed and discussed in Sect. 3.3 and is now evidenced in the FRF in Fig. 7. The boost observed for the frequencies around 600, 1000 and 2400 Hz, in Fig. 5, can be better understood now, when looking at the FRF in Fig. 7. The first three resonances, corresponding to the modes of the membrane, are very close to the aforementioned frequency values.

Not only the resonances, but also the band-pass filtering that the soundbox applies can be observed in Fig. 7, specially the sharp slope in the high-cut edge, around 3000 Hz. Later gramophone models could already profit from the development of the electrical equivalent circuit theory [14]. A better impedance matching between the different parts was possible, extending this frequency range and softening the high-cut. But this is not the case for the gramophones that were on the market in the first decades of this technology, as the here investigated model.

4.2 Influence of the Tonearm

The tonearm is the element of the reproducing chain that is equivalent to the tube in the recording chain (Sect. 3.2). When attaching a tonearm to the soundbox, the different ending impedance causes the shifting of the previously observed membrane resonances and adds tube resonances, as evidenced by the comparison between the grey (identical to the plot in Fig. 7) and black lines in Fig. 8.

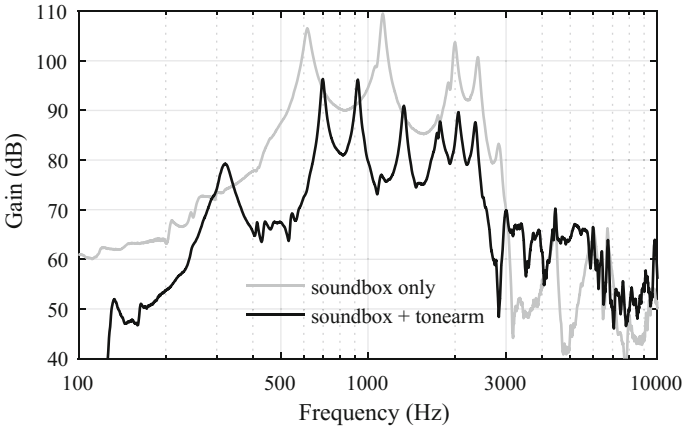


Fig. 8 FRFs of soundbox and tonearm

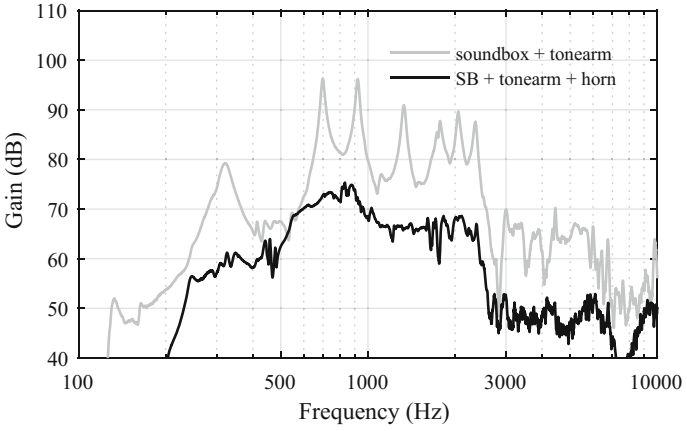


Fig. 9 FRFs of the academy gramophone

4.3 Influence of the Horn

Differently from the recording horns, that are predominantly conical, the reproducing horns gradually assumed an exponential shape. The result is a better impedance matching with the surrounding air, enhancing the sound radiation and providing a flatter frequency response function. These behavior can be observed in the FRFs in Fig. 9, when comparing the grey (previous setup, with soundbox and tonearm, but without horn) and black lines. When attaching the horn, the FRF becomes much flatter, since the enhanced radiation means also that there is less energy available for building up the sharper resonances that were observable in the previous setups.

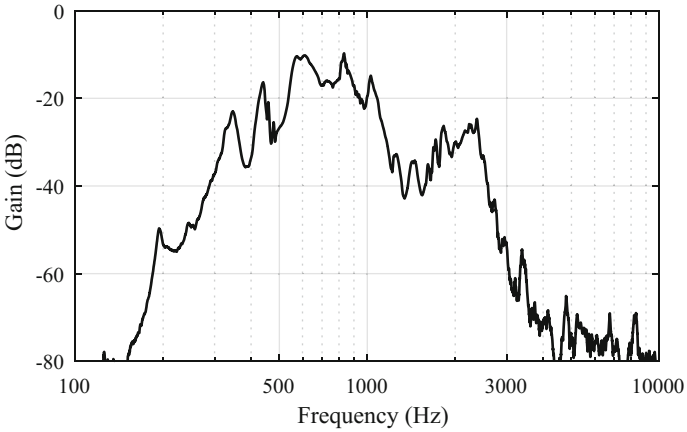


Fig. 10 FRF resulting from the multiplication of recording and reproducing FRFs

One aspect that is still very present in the FRF of the complete setup (black line in Fig. 9) is the sharp slope in the high-cut edge, already discussed in Sect. 4.1.

5 The Combined Transfer Path

It is now possible, based on the FRFs obtained for the recording and reproducing setups (black lines in Figs. 5 and 9, respectively), to analyze their combined influence[9]. This can be achieved by multiplying both spectra, resulting in the frequency response shown in Fig. 10.

It can be observed that only some of the sharp resonances in the recording FRF (black line in Fig. 5) become less prominent when multiplying both spectra, like the ones around 200 and 600 Hz. Most resonances and slopes (like the ones around 1 and 3 kHz) introduced by elements of the recording setup, however, remain quite sharp in the combined FRF. They dominate the overall frequency response, while the influence of the here investigated gramophone can, in general terms, be limited to a slight boost in the range between 500 and 1000 Hz and a band-pass characteristic with low and high-cut frequencies around 250 and 2500 Hz (see the black line in Fig. 9).

The results shown and discussed in Sects. 3 and 4, and in this section are based on measurements performed on specific recording and reproducing setups. Thus, only general aspects of the phenomena here discussed should be extended to other setups. The dominance of the recording setup resonances in the combined FRF, for

[9]The term "combined" transfer path is intentionally used instead of "complete", since there are aspects like the wax impedance and the influence of using different needles that were neglected in this work, as mentioned earlier in the text.

instance. But the individual resonances will have different central frequencies and damping if using another recording setup or when the media is played on a different gramophone model.

6 Interaction of Voice and Device

Whereas in modern voice recordings the professional devices used for recording and reproduction of voice are designed to induce no or negligible artefacts and only apply intended modifications to the voice signals, historic devices used for professional recordings and reproduction of music significantly changed the signals of the musician in front of the recording horn in many ways.

The most obvious impact of the signal chain is the presence of noise and the band limitation of music signals. The noise is mainly produced during the mechanical engravement process but also results from the overall low dynamics of the signal paths described in the previous sections. This limitation of the lower end of the signal dynamics results in masking effects of subtle musical sounds as present in soft or whispered voice. The upper dynamical limit is given by the mechanical transformation of sound to the groove on cylinder or disc. Since the elongation of the needle cannot be larger than the distance between two neighboured grooves, any larger amplitude in lateral recordings would either be limited/clipped or destroy the recording. Also, the force needed to engrave was limited due to the purely mechanical function of the soundbox.

The soundbox not only limits the dynamics of the recordings during the recording and reproduction process but also introduces strong deviations from a flat pressure-to-elongation transfer function, as the measurements presented in Sects. 3.3 and 4.1 showed. An observation of the membrane vibration exhibits the presence of modes that strongly affect the movement of the needle in the frequency range of the voice signals [15]. In both engravement methods—vertical and lateral—the presence of membrane modes in the frequency range of the voice signals would lead to reduction and increase of amplitudes for those frequencies in the voice signal that match the eigenfrequencies of the membrane modes.

The recording and reproduction horns as well as the tubelets that are used to connect horn and soundbox or several horns during recording of ensembles, add further limitations and modifications to the overall transfer function as described in the previous sections. Consequently, the recorded singing voice that reaches the listeners' ears represents a strongly modified voice signal.

The possible influence of attributes of devices from the mechano-acoustical era on instrumental musical practice has been a topic in a few works over the last two decades; more specifically, the changes in the portamento technique of violin playing [8, 9]. In the last years only, this discussion has started to be extended to possible influences on the singing technique development [13].

Since the interpretation of voice signals makes use of cues such as formant frequencies, amplitudes, and harmonics-to-noise ratio to interpret the performance and

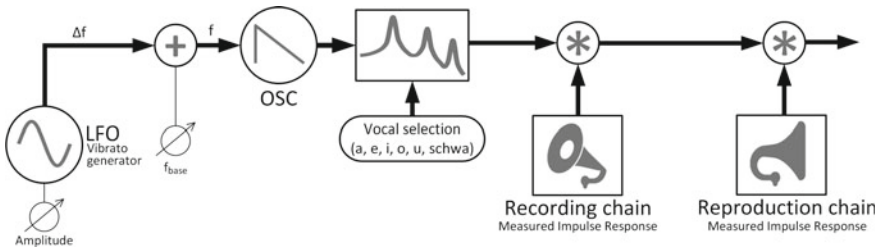


Fig. 11 FRF resulting from the multiplication of recording and reproducing FRFs (from [10])

style of singers, the modifications that early recording devices introduce may lead to an interpretation of recordings that differ from those that a modern recording would provide. As an example, the modifications of the voice formants due to the transfer function could be interpreted as articulatory expression that never was intended or performed by the singer. An ornament such as vibrato or glide could be enhanced or reduced by the presence of resonances in the transfer functions without any intention of the singer. Whereas these effects can be predicted and also be heard in recordings, they are not easily reproduced due to the lack of historic recording devices.

The frequency response functions of a recording setup, a gramophone, and their parts were the basis for the discussions in the last three sections. It was possible to link some resonances to specific parts of the devices and observe the interaction between them when connecting the parts. The FRFs of these and of other devices can also be useful for investigating the interaction of voice or other sound sources with attributes of the devices. This can be implemented by convolving dry voice recordings or synthesized voice with the impulse response[10] of a specific recording and/or reproducing setup.

A systematic investigation of the impact of specific parameters of the voice and the device parts can be achieved by using an approach that combines modeled and measured acoustic properties of the complete signal path. In Fig. 11, the block diagram of such a voice-device interaction model is shown. A voice signal is modeled using a multi-sine signal with a fundamental frequency that can be characterised by its average pitch and the ambitus and frequency of its periodic variation (vibrato). The voice signal is convolved with arbitrary transfer functions of the device to evaluate the interactive effect of singing voice parameters and the acoustic device properties. First investigations using a modelling approach indicate that such modifications can be expected with amplitude modulations of more than 10 dB for selected sounds [10].

[10]The impulse response is the time-domain equivalent of the FRF.

7 Discussion

The investigations of historic devices' FRFs and the comparison of voice signals at both ends of the production chain demonstrate that early devices used for recording and reproduction can significantly change the timbre, ornaments and formants of voice signals. The results show that recordings performed with mechanic-acoustical setups could never have a flat frequency slope, as would be desired for documenting musical practice for ethnomusicological studies in a "neutral" way.

According to first investigations using a modeling approach, a linear approach seems to be sufficient to explain the reduction or amplification of timbre changes and formant shifts. The effects are strongly dependent on the individual formant characteristics as well as on the specific use of ornaments such as vibrato and glides. Also, the effects do not occur for all notes in a musical piece which matches well the results from simulations of the voice-device interaction.

Masking, together with bandwidth limitations can change the perceived voice from bright to flute-like or even falsetto. The change of voice character can therefore be partially attributed to the signal modification caused by the recording and reproduction devices.

Future work could extend the modeling approach using convolution with forward and inverse FRF paths. This concept could provide sound signals that add the modifications of historic devices to modern recordings or reduce the influence of the devices from historic recordings.

An extended version of the simulation program could provide the user with an interface that allows the selection of historic device parts and then generates the corresponding voice sound. This concept could be used to directly compare the effect of the choice on the sound of voice recordings. This would contribute to the understanding of device development and the choice of adequate parts for recording tasks.

The last step would be the attempt to interpret historic recordings in the light of the knowledge and by using simulations to understand the implication of historic devices on the performance practice of singers. The overall goal would be to learn to what extend the limitations and characteristics of the recording conditions influenced the choice of singers, pieces, style and ornamentation.

Acknowledgements The authors would like to thank the German Research Foundation (DFG—project number 289601849) for funding this work. As well as the Detmold University of Music and the Paderborn University for the institutional support. Our deep gratitude to George Brock-Nannestad for lending the recording horn replica and for technical advice. The singers Doris Maria Ritter and Boris A. Bolles are acknowledged for the contributions of their voices to the investigations.

References

1. Brock-Nannestad G (1981) Zur Entwicklung einer Quellenkritik bei Schallplattenaufnahmen. Musica 35:76–81
2. Brock-Nannestad G (1997) The objective basis for the production of high quality transfers from pre-1925 sound recordings. In: Audio engineering society convention, vol 103
3. Copeland P (2008) Manual of analogue sound restoration techniques. The British Library
4. Dromey C, Carter N, Hopkin A (2003) Vibrato rate adjustment. J Voice 17(2):168–178
5. Hähnel T (2018) Über die Quantifizierung des Heldentenors. Vibrato, Ornamentik, Glissando, Tempo und Register in akustischen Tonaufnahmen zwischen 1900 und 1925. In: Beiträge zur Tagung der Gesellschaft für Musikforschung. Oldenburg (2018) (Accepted talk and paper)
6. Ifukube T, Kawashima T, Asakura T (1989) New methods of sound reproduction from old wax phonograph cylinders. JASA 85(4):1759–1766
7. Iwai T, Asakura T, Ifukube T, Kawashima T (1986) Reproduction of sound from old wax phonograph cylinders using the laser-beam reflection method. Appl Opt 25(5):597–604
8. Katz M (1999) The phonograph effect: the influence of recording on listener, performer, composer, 1900–1940. PhD thesis, University of Michigan
9. Katz M (2006) Portamento and the phonograph effect. J Musicol Res 25(3–4):211–232
10. Kob M, Amengual Garí SV, Bolles BA, Ritter DM, Pirch P (2018) Influence of early recording and playing devices on voice sounds: modification of singing voice formants. In: Fortschritte der Akustik - DAGA 2018, pp 1703–1706
11. Kob M, Henrich N, Herzel H, Howard D, Tokuda I, Wolfe J (2011) Analysing and understanding the singing voice: recent progress and open questions. Curr Bioinform 362–374
12. Kolkowski A, Miller D, Blier-Carruthers A (2015) The art and science of acoustic recording: re-enacting Arthur Nikisch and the Berlin Philharmonic Orchestra's landmark 1913 recording of Beethoven's Fifth Symphony. Sci Museum Group J
13. Martensen K, Zakharchuk P, Kob M, Grotjahn R (2015) Phonograph und Gesangsstimme: Untersuchungen zur Akustik früher Aufzeichnungs- und Abspielgeräte. In: Fortschritte der Akustik - DAGA 2015, pp 1429–1432
14. Maxfield JP, Harrison HC (1926) Methods of high quality recording and reproducing of music and speech based on telephone research. Bell Syst Tech J 5(3):493–523
15. Pirch P (2015) On the use and acoustical characteristics of mechanically augmented musical instruments and early recording devices. Master's thesis, Detmold University of Music
16. Sagers JD, McNeese AR, Lenhart RD, Wilson PS (2012) Analysis of a homemade Edison tinfoil phonograph. JASA 132(4):2173–2183
17. Sundberg J (1974) Articulatory interpretation of the singing formant. J Acoust Soc Am 55:838–844
18. Weege TA, Habasinska D, Kob M (2018) Influence of early recording and playing devices on musical sound: FRF measurements of horn, soundbox and tonearm. In: Fortschritte der Akustik - DAGA 2018, pp 1707–1710
19. Welch W, Burt L (1994) From tinfoil to stereo: the acoustic years of the recording industry, 1877–1929. University Press of Florida
20. Wilson P, Webb GW (1929) Modern gramophones and electrical reproducers. Cassel and Company, Ltd