

Change and Continuity in Sound Analysis: A Review of Concepts in Regard to Musical Acoustics, Music Perception, and Transcription

Albrecht Schneider

1 On measuring Basic Properties of Sound: A Brief Retrospect

Over the past decades, a broad range of software and hardware tools has become available suited to perform sound analysis both in the time domain and in the frequency domain. Though musical acoustics as a field of research based on both calculation and experiment started to evolve around 1600 (cf. Cohen 1984), and had gained impetus by 1700 with experiments on, for example, vibration of strings as conducted by J. Sauveur and many other scholars since (see Cannon and Dostrovsky 1982), investigation of actual sound as produced by instruments and voices was limited since appropriate tools for measurement and analysis were scarce. Experiments at the time of Chladni (1805) were mostly done with sets of tuning forks or organ pipes (see Beyer 1999). After Charles Cagniard de la Tour had invented the siren (1819), such instruments came into use soon (as in the pioneering experiments of August Seebeck that led to the first formulation of ‘periodicity pitch’ published in 1841; see Hesse 1972, 58ff.; de Boer 1976; Schneider 1997b, 134f.). Tuning forks, pipes, sirens as well as resonance boxes and resonance bottles such as described by Helmholtz (1863/1870/1896) were the basic toolkit of the acoustician then. Even Stumpf conducted most of his experiments on perception of consonant and dissonant sounds (1890, 1898) with the aid of sets of tuning forks (mounted on resonance boxes), reed pipes and organ mixture stops. One proven method used to study the structure of harmonic partials in complex tones was resonance, another was additive synthesis of sounds such as

This article is dedicated to the memory of Walter Graf (1903–1982), one of the pioneers of sound research in Systematic and Comparative Musicology.

A. Schneider (✉)
Institut für Musikwissenschaft, Universität Hamburg,
Neue Rabenstr. 13, D-20354 Hamburg, Germany
e-mail: aschneid@uni-hamburg.de

vowels by means of a set of tuning forks that could be excited in an electromagnetic circuit, and thereby would produce a continuous sound. Mechanical devices such as the *Tonmesser* built by Appun (of Hanau, Germany) and the *Tonvariator* developed by Stern (1902) even allowed measurement and production of tones varying continuously in frequency as well as synthesis of sonorities ranging from perfect harmonic to inharmonic. Besides resonance, interference was a principle used for analysis. By the end of the 19th century, an apparatus for dampening or cancelling partials out of complex sounds (such as vowels) had been developed (cf. Graf 1980, 212). A range of mechano-acoustical devices available at the time were used by Helmholtz and other scholars (e.g., Carl Stumpf and co-workers, Wilhelm Wundt and co-workers) to explore sound as well as auditory phenomena such as the sensation of consonance and dissonance, roughness, beats, and combination tones.

For the study of sound wave characteristics, Helmholtz (1870, 33f.) also saw the need to record sound as radiated from musical instruments (including the human voice) on some suited graph paper or other material. He rightly emphasized that the shape of a sound wave per period determines the corresponding timbre (German: *Klangfarbe*; see below, Sect. 3). Helmholtz pointed to the *Phonautograph* of Scott de Martinville who, around 1857–1860, had actually recorded sound waves (including a few seconds of a French folk song, *Au clair de la lune*) on a rotating cylinder.¹ In 1877, Edison presented his *Phonograph* (the early tinfoil version), which was used in the 1880s and 1890s for the study of sounds from musical instrument and the human voice by the German physiologist Georg Meissner (see below). With Edison's improved 1888 model of the *Phonograph* (Edison 1888), recording and reproduction of sound had become a standard method. In particular the physiologist Ludimar Hermann published numerous articles on speech sounds (vowels, consonants; see below, Sect. 3) recorded and analyzed by way of *Phonophotographie* (e.g., Hermann 1889, 1893, 1894, 1895, 1911). As an alternative to the *Phonograph*, the physiologist Victor Hensen had developed a machine called *Sprachzeichner* (1879, 1888) that offered a very subtle registration of sound waves (plus the recording of a tuning fork as referent). The *Sprachzeichner* was one of the fundamental tools in phonetics that has been employed, for instance, by the Finish-Swedish linguist, Hugo Pipping, for the study of spoken and sung vowels (e.g., Pipping 1894). The study of sung vowels of course played a central role in the theory of formants (see below, Sect. 3.3). Meissner, Hermann, Pipping and others took great pains in analyzing complex sounds for finding the period length and the fundamental frequency as well as to determine the spectral content for each single period. Fourier analysis was done by

¹ It is reported (cf. Beyer 1999, 140f., Figures 6,7,8,9,10,11) that Karl Rudolph Koenig (who invented a range of instruments for acoustics) had improved the *Phonautograph* by adding a conical horn to collect the sound (as was similarly done later by Edison with the *Phonograph*) as well as the rotating cylinder on which the sound was actually recorded as an oscillogram. As to the early history of sound analysis, see also Graf (1980, 211ff).

hand with the aid of rulers and curve templates in a time-consuming geometrical analysis procedure.²

2 Frequency Measurement, Periodicity Estimation, Melographic ‘Pitch’ Notation

Besides Meissner, Hermann, and Pipping, another scholar with a medical training, Edward W. Scripture, became a specialist in the study of speech curves for which he developed a methodology to determine actual (fundamental) frequencies of speech or sung melodies (Scripture 1906, 1927). This approach still was based on a manual reading of wave-lengths (Scripture 1906, 60ff.) or periods that had to be transferred to their respective frequency values. A similar yet improved technique again called *Phonophotography* was employed by Metfessel (1928) who had the sound wave (plus a referent vibration of 100 Hz obtained from a tuning fork usable as a time-line) recorded on film. It was applied to the study of vibrato and intonation in “western” music as well as to the microstructure of African-American singing styles.

The basic idea behind the analyses performed by Hermann, Scripture, Metfessel as well as by other researchers is to use the information provided by the temporal structure of a sound wave in order to determine the frequency (Hz) corresponding to a given period length (ms). The fundamental frequency can be calculated by making use of the inverse relation between period T (ms) and frequency f (Hz), which is given by $T = 1/f$, and $f = 1/T$. In case the signal is a complex tone comprising a series of harmonic partials, the period typically is determined by the first partial acting as the fundamental frequency f_1 , which is acoustically real if it can be measured as a spectral component (for example, in the sound of a flute in normal blowing condition). In the case of a harmonic complex tone, partial frequencies can be determined according to $f_n = n \times f_1$. In a strict sense, the frequency corresponding to any particular period of a sine tone or to f_1 of a complex tone (with all partials locked in zero-phase) can be taken as its instantaneous frequency (German: *Augenblicksfrequenz*).³ An additional aspect is that, in a signal where f_1 is weak or even missing, the period length T (ms) determined by a sufficient number of harmonic partials superimposed on each other and locked in phase is identical with the period length τ (ms) defined by the f_1 partial (for

² Some of the early research is reported (with details in regard to methods of measurement and the technology then available) in Meissner (1907), Krueger (1907), Hermann (1893), Herrmann (1908). See also Panconcelli-Calzia (1941/1994).

³ This is a practical definition applicable to empirical observation and measurement. In signal processing, the notion ‘instantaneous frequency’ has a more technical definition based on calculating the instantaneous (or local) phase $\varphi(t)$ (a real-valued function) for a (complex-valued) function $x(t)$ representing the signal; the instantaneous angular frequency can be determined as the derivative of the phase, that is $\omega(t) = \varphi'(t)$; see Cohen (1995, 39ff).

examples, see Schneider 1997a, b and 2000). What is measured, in this case, is the repetition frequency of the envelope of a complex waveshape. This repetition frequency can be labeled f_0 or F_0 , and in contemporary psychoacoustics is generally addressed as the “fundamental frequency” of a complex sound (what is rather misleading since it implies the concept of a complete harmonic spectrum where f_1 is present). The envelope repetition frequency f_0 of course permits to calculate also f_1 (whether this is present in the spectrum or not; if it is, $f_0[\text{Hz}] = f_1[\text{Hz}]$). The sensation corresponding to f_0 is a ‘periodicity pitch’ that proved of importance to explain pitch perception in the case of a ‘missing fundamental’, that is, where the sound carries little or no energy at f_1 (cf. de Boer 1976). The sensation of a ‘periodicity pitch’ evoked from detecting f_0 at times has also been labeled ‘low pitch’ since it is typically found below the spectral components actually present. For example, in major and minor chords in “root” position comprising a triad of three complex tones each synthesized from a number of partials in just intonation, the “fundamental” f_0 that can be calculated by either autocorrelation or subharmonic matching appears as the common denominator of the total harmonic series (cf. Schneider and Frieler 2009; Schneider 2011).

In practice, finding the frequencies for musical tones from reading periods manually (as was done by Scripture et al.) is an arduous task since the sound segments under investigation must be small to allow manual search for appropriate zero crossings defining periods. By about 1930, measurements had been improved in that electronic equipment including amplifiers and oscilloscopes (based on either galvanometer or cathode ray technology) had become available that allowed for continuous recording of oscillograms up to ca. 15 s on film or special paper strips. Such a setup has been used, for example, to study intonation patterns of cellists playing scales and melodic phrases in a microtonal context (Kreichgauer 1932). Still the length of periods for each tone played by skilled musicians had to be determined manually (i.e., by counting the number of periods per time unit); the amount of labour Kreichgauer (1932) has put into his study of intonation patterns seems incredible since a corpus of several hundred meters of oscillographic registrations resulting from a series of experiments had to be analyzed.

The advantage of taking the sound wave as a prime source for analysis is that, if we assume linear behaviour of all tools used in the recording, it can be assumed that no alterations in the signal have yet occurred before analysis. Hence, looking for the period lengths of a time signal and transferring the temporal information to corresponding frequency values can be regarded as an objective way of measurement. Such an analysis allows, most of all, to determine the fundamental frequency per period, and to see if, and possibly to which extent, there is a fluctuation in period length (meaning that there is some frequency modulation in the signal as well). To illustrate the case, we might look into a very short segment of the sound wave recorded as part of a song, *Abu Zeluf*, as sung by a woman in Lebanon (Fig. 1).⁴

⁴ *Abu Zeluf*, sung by Dunya Yunis, recorded by Poul Rovsing Olsen on 4th of February, Beirut 1972. Published on the LP *Music in the World of Islam, Vol. 1: The Human Voice*, Tangent

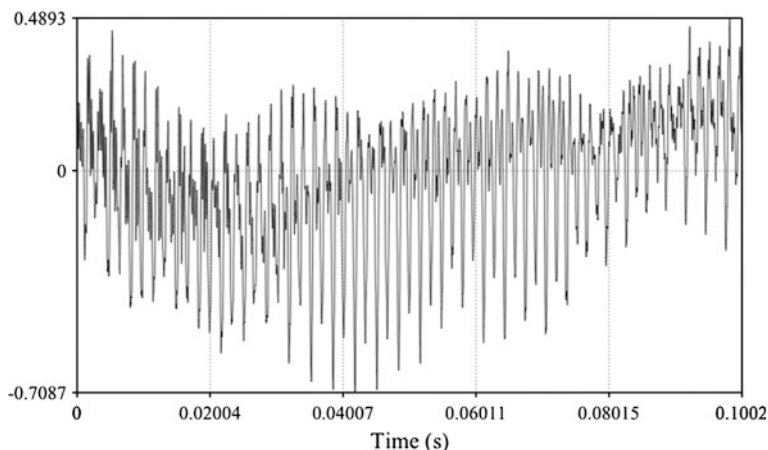


Fig. 1 Short segment (100 ms) of sound recorded from a female singer in Lebanon

The period length at the beginning of this segment is 3.426 ms corresponding to 291.91 Hz while at the end a period length of 3.025 ms is found corresponding to 330.54 Hz. Hence, there is a shift in fundamental frequency that can be viewed also as a shift in ‘pitch’ (in regard to hearing and psychoacoustics; see below) by more than one semitone. With modern digital equipment, measuring the period lengths and even calculating the ‘pitch shift’ can be done with great precision. In the days of analogue measurement, tracking of zero crossings that define the periods in complex wave shapes mostly was done with an oscilloscope, and often the signal was put through a low pass or band pass filter in order to simplify the wave shape, and possibly to extract the fundamental from each complex tone (as this was believed to represent the ‘pitch’) in a melody. Already in the 1930s, techniques for finding the fundamental frequency of a sequence of tones forming a (quasi-continuous) melodic contour and plotting such a curve over time by means of a so-called *Tonhörensreiber* or *pitch recorder* suited for monophonic signals had been developed (Grützmaier and Lottermoser 1937; Obata and Kobayashi 1937). The range of the German *Tonhörensreiber* has been given as 2.5 octaves, and the precision of ‘pitch’ as ca. 25 cent (Lottermoser 1976/1977, 139). By the end of the 1940s, ideas of Charles Seeger, one of the pioneers in systematic and comparative musicology, for constructing a ‘melograph’ capable to plot the ups and downs of a melodic contour as well as the dynamic changes of a signal over time had reached the state of a beta version (cf. Seeger 1951). The aim was to develop an electronic device that could perform *electronic sound-writing in the laboratory* (Seeger 1951), that is, automated transcription of (monophonic) music recorded in the field or available on phonograph records. Melographic analysis was

(Footnote 4 continued)

Records TGS 131 (London 1976), B2. As to the history of the recording and its later (ab)use, see Feld and Kirgegaard (2010).

considered not only to make transcription easier and subtle changes in pitch detectable but also to provide a tool for objective analysis and notation independent of the ear of the individual listener (see also Schneider 1986, 1987).

During the 1950s, more attempts at constructing ‘melographs’ were made (of which one realized in Oslo was notable; see Dahlback 1958). Seeger’s idea for a melograph finally materialized in his ‘Model C’, which was used as an aid in transcription of, for example, saxophone parts Charlie Parker had played in his rendition of *Parker’s Mood* (rec. New York 1948; cf. Owen 1974). Model C besides tracking of ‘pitch’ offered recording of dynamics as well as a (rather elementary) representation of spectral energy over time. A more advanced concept of a melograph was developed by Miroslav Filip who, instead of low-pass filtering a given signal for extraction of the fundamental (f_1 of a complex tone) took a nonlinear approach to envelope periodicity detection that could handle also signals with a “missing fundamental” and proved to be robust in actual measurement (Filip 1969, 1970). The melograph developed by Filip has been used as an aid for transcription and analysis of orally transmitted music (cf. Elschek 1979, 2006). The main purpose of melographic analysis is to determine the exact fundamental frequency contour or (for harmonic/periodic signals) its equivalent, the ‘pitch contour’ calculated from f_0 values.⁵ Also, the onset and duration of tones as they occur in a certain melodic context are of interest. In addition to finding pitch contours at large, one may look into details of intonation and fluctuations of pitch due to glissandi, vibrati, etc. An example to illustrate the case is given in Fig. 2, which shows the ‘pitch’ trajectory of a segment of ca. 20 s of a female singing a song, *Abu Zeluf*, recorded in the Lebanon.

The segment in question includes three short introductory notes (the musical notes being, roughly, c4 at about 264 Hz [repeated once], and d4 taken a bit sharp at ca. 301 Hz) followed by a very long note e4 (at ca. 333 Hz) which is held almost constant in pitch for more than 9 s. Then, after a short e4 and a sudden jump upwards to a short peak at ca. 380 Hz, a strong melisma follows with a modulation between the notes e4 and d4. Of course, all notes or, rather, note names like c4, d4, e4, etc. must be taken as relative pitches only that have to be interpreted in regard to the Arab modal scale used in this song (apparently, a scale of the *maqam rast* group of modes). The modulation frequency found in the melisma is about 5.5 Hz (Fig. 3):

After the melisma, we find a sequence of short notes beginning on b3 (at ca. 245 Hz), with a transition from c#4 and d4 to e4, followed by a d4, another e4 and then d4 (on which note the phrase ends). What the melogram (Fig. 2) clearly reveals is the small fluctuation of ‘pitch’ on the long note e4 before the melisma, and the fairly regular modulation applied when rendering the melisma $d4 \leftrightarrow e4$.

⁵ The measurements for pitch curves shown in Figs. 2 and 3 were performed with the Praat software (Boersma and Weenink 2011). A special autocorrelation method (Boersma 1993) was chosen for calculation of pitch with a time resolution of 1 ms.

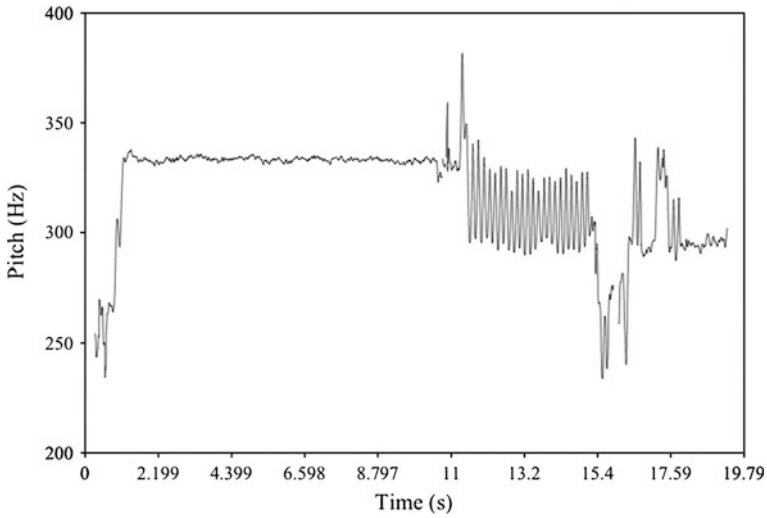


Fig. 2 Melogram of a segment (ca. 20'') of a female singing *Abu Zeluf*, Lebanon

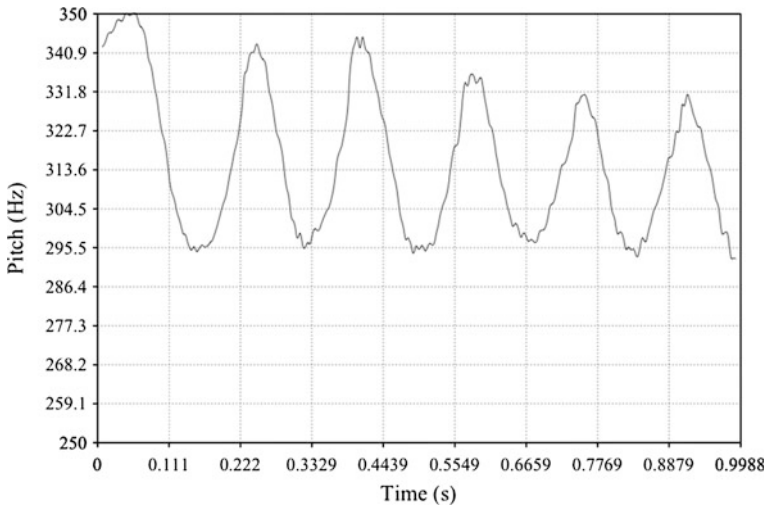


Fig. 3 Dunya Yunis singing *Abu Zeluf*, melisma, frequency modulation

In addition to continuous registration of the fundamental frequency as a function of time, another approach was taken by establishing histograms of fundamental frequency distributions (Tjernlund et al. 1972; Filip 1978). This approach includes statistical considerations and needs computers for practical realization. Recently, such a histogram analysis was chosen (in modern software implementations) for the study of Turkish *makam* music (Bozkurt 2008; Bozkurt et al. 2009;

f0 calculation is done with the well-known *Yin* algorithm of de Cheveigné and Kawahara (2002) and of Cambodian melismatic chanting (of the genre *smot*; Bader 2011). In these recent studies, melographic analysis is combined with an automated, algorithmic analysis of fundamental frequency distributions. To be sure, while melographic analysis can show deviations of ‘pitch’ from expected values (i.e., frequency values defined by a certain tone system or scale) that occur at a certain time of a performance, and in a specific melodic context, frequency histograms eliminate the time dimension as well as the melodic context in which certain pitch deviations or melodic embellishments (such as vibrato or melisma) take place. The advantage of the histogram technique, however is that the number of occurrences of particular fundamental frequencies leads to a pattern of pitches prevalent in a certain piece of music. In most cases (depending of course on the shape of the frequency distribution and the statistics for frequencies obtained in a particular study) a melodic scale or mode can be inferred from the histogram data.

The melographic approach that, in its original concepts, was directed to finding trajectories of fundamental frequency which then could be plotted as pitch contour curves against time on some graph paper (possibly with a linear or logarithmic frequency [y] and a linear time [x] scale), has been expanded in several directions over the past decades. First, besides normal algorithms suited to achieve either f_1 (= fundamental frequency of complex tones/sounds with n harmonics; f_1 is typically extracted by low pass or band pass filtering) or f_0 (= frequency with which the envelope of a complex sound repeats per second), a number of ‘tracking algorithms’ have become available that allow to calculate trajectories for partials according to some model. One of the best-known such models is that issued by McAulay and Quatieri (1986) that utilizes frame-to-frame peak matching, thereby establishing a number of frequency tracks (depending on the spectral energy distribution and the number and strength of peaks per frame). Instead of covering a full spectrum, tracking algorithms may be directed to the fundamental frequency of complex sounds determined by means of, for example, a constant Q transform (Brown and Puckette 1993; see also Brown 2007). This approach allows precise tracking even of signals changing rapidly in frequency over time (like the example shown in Fig. 1). Other methods applied to speech and music for f_0 estimates include autocorrelation plus additional processing steps (de Cheveigné and Kawahara 2002) and Principal Component (PC) autoregressive frequency estimation based on the ModCov (modified covariance, see Marple 1987) model (Hekland 2001). Also, software especially suited for the study of micromelodic ornamentation (such as found in Hindustani music of India) has been developed that uses autocorrelation for an initial pitch estimate but also performs calculation of the spectral centroid; smoothed pitch contours are found by fitting Bézier spline curves to the data (Battay 2004).

Second, while algorithms suited to extract either f_1 or f_0 from complex sound for a long time were restricted to deal with monophonic signals only (i.e., signals containing one tone and, typically, one pitch per time), there have been a number of attempts in recent years to cover polyphonic signals as well (for an overview, see Klapuri 2004; Klapuri and Davy 2006, part III, chs. 7 and 8). The goal of

polyphonic analysis often is to provide (if possible) a MIDI file of musically defined (pitch, duration) notes from audio sample data (see, e.g., Paiva et al. 2008). Some of the concepts chosen are related to those of Auditory Scene Analysis (ASA, Bregman 1990) since a number of source signals have to be separated into (spectrally correlated) “streams” or, in this case, “voices”. However, though a range of commercial and free software for PCM-to-MIDI-conversion is at hand, one has to bear in mind that the rate of accuracy achieved for polyphonic pitch tracking analysis even with advanced models by now comes close to 60 % (cf. Cañadas Quesada et al. 2010). As it seems, there is still a lot of work ahead for automated transcription of polyphonic music.

3 ‘Sound Colour’ (*Klangfarbe*) and Spectrum: Acoustical and Psychological Aspects

3.1 *On the Archaeology of ‘Sound Colour’ or ‘Timbre’*

Man (like other mammalian species) has experienced various types of sounds as occurring in the natural environment for thousands of years. Also, communication between humans as well as with animals needed articulated sound. It is very likely that individuals discovered phenomenal differences as well as similarities between various sounds early on, and that categorization of sounds according to certain attributes was undertaken. Though direct evidence is difficult to adduce, there are “ethnographic parallels” that at least can illustrate the case (see, e.g. Feld 1990 with a survey of sound phenomena that are part of the natural environment as well as of the sociocultural life of the Kaluli tribe of Papua New Guinea). Regarding ‘Old World’ cultures, there is ample evidence for sound phenomena resulting from singing or from various instruments in written sources of Greek and Roman antiquity, and also in medieval writings. For example, the mathematician and music theorist Nikomachos, dealing with sounds produced by various instruments (harm. en. IV), rightly states that sound is an impact on air, and that sounds can be distinguished as to large and small, dull and sharp (low pitches corresponding to dull sound because of a loss of tension in stringed instruments). By the end of the Middle Ages, and more so during ‘Renaissance’, musical instruments in use had increased by far, and so had the number of different stops or ranks of organ pipes producing different sounds (the variety of instruments is reflected in Praetorius’ *De Organographia* of 1619). These facts indicate an awareness of people listening to music for characteristic tonal registers and sound qualities.

As is well-known, theory of vibration and other fundamental issues in acoustics were developed steadily between, roughly, 1500 and 1800 (cf. Cannon and Dostrovsky 1982; Cohen 1984). Resonance in strings and stringed instruments was a major topic of research already when Sauveur (1701) succeeded in the determination of harmonic partials in vibrating strings. However, his lectures, demonstrating that

even a single musical ‘tone’, when played on, for example, a harpsichord, contained a series of harmonics, were widely recognized, and had great impact also on music theory (as is evident in particular in Rameau’s writings, see Schneider 2011). Since vibration theory was based on observations of the swinging pendulum, the notion of a period of vibration (expressed as the duration the pendulum needed to complete a full circle of motion) and its relation to ‘pitch’ was fairly well understood. The faster the motion, the higher the frequency of vibration and the number of pulses transmitted through the air to the ear would be. And the stronger the excitation force applied to a string, the louder the resulting sound. ‘Pitch’, then, was dependent on the number of pulses per time unit, and ‘loudness’ on the amplitude of vibration. Also, one did distinguish ‘tone’ (the result of a regular vibration) and ‘noise’ (irregular or even arbitrarily changing motion).

When Chladni entered the stage with his comprehensive work *Die Akustik* (1805/1830), he rightly argued that the source of audible sound (German: *Schall*) is an elastic body set to regular vibration (*Klang*) or irregular motion (*Geräusch*). Criticising Rameau and others who had claimed one could hear a number of harmonic partials in addition to the fundamental (Chladni 1805, 3, in this respect, speaks of *Hauptschwingung*), Chladni argued that a given sound is not an agglomerate or complex but something that is single (*etwas ganz Einfaches*). Notwithstanding his many insights into types of vibration (transversal, longitudinal, torsional) and the modal structure of vibrations in elastic bodies (many of them demonstrable by means of *Klangfiguren*), Chladni did not quite come to grips with ‘sound colour’ or *timbre* though he mentions this French term in a chapter treating modifications and articulations of sound (Chladni 1805/1830, § 248). In line with his fundamental theory of sound as emanating from vibration of elastic bodies, he saw such modifications and articulations of sound depending on various materials (such as steel, brass, or gut used for strings) as well as in small differences in the motion within vibrating bodies resulting probably from strain and stress. This was a modern view as far as vibration is concerned yet not enough in regard to the perception of ‘sound colour’. In a later writing, Chladni (1817, 58) again refers to the French word *timbre*,⁶ saying that it denotes *die qualitative Verschiedenheit des Klanges auf die Wirkung, wofür man im Deutschen keinen bestimmten Ausdruck hat*. This indicates that *timbre*, or *Klangfarbe* by this time was not yet a central issue for research. Moreover, Chladni’s resistance against Rameau’s perception of string partials as audible tones led even major music theorists in Germany to subscribe to his view. For instance, Weber (1817/1824/1830, §§ I–V, 180) relates to Chladni when stating that a *Klang ist ein einfacher und ungemischter Laut*, and that the very nature of musical instruments is to produce sounds or tones (a *Klang* of a definite pitch is called *Ton*) as pure as possible. Only instruments of a musically “lower” grade (like a snare drum and, more so, cymbals and other Turkish instruments) apparently have a number of

⁶ The meaning of *timbre* in the French language comprises sound, sound colour, and also bike bell (*timbre d’une bicyclette*) as well as brand, mark and (postal) stamp.

Nebenschwingungen that result in audible *Beitöne*. These, however, Weber claims, are almost not audible in “higher” instruments such as strings and wind instruments. Partials, he (1830, 12) concludes, even though these might be found in vibrating strings, have nothing to do with the nature or the beauty of a sound and can be considered an imperfection that, however, is harmless since higher partials or *Beitöne* in reality would not be audible. Such statement left Weber in a position where he, on the one hand, could not but admit that there is something one perceives as the quality or character of a sound (*sein eigenthümliches Gepräge* in the sense of the French *timbre*), and on the other he refuses to take notice of the acoustical basis underlying any *Tonfarbe* or *Klangfarbe*.

Due to ongoing research in acoustics and also in optics (where the wavelength and spectrum of light had become an issue since Newton, and, by analogy, sound thus could hardly be conceived without the concept of a spectrum), the situation had changed by about 1850. In Opelt’s *Theorie der Musik* (1852, § 7) it is hypothesized that “the so-called *Klangfarbe* depends on different kinds and shapes of pulses” (by pulses periodic wavetrains are meant), even though this cannot be proved with certainty by ear alone. Opelt assumed that sound as transmitted from an instrument indeed is a complex whole to which several vibrating structures of each instrument contribute.

Since ‘pitch’ was associated with wavelength and the number of vibrations per time unit, and the sensation of intensity (or, rather, loudness) with the amplitude of vibration, ‘sound colour’ could not be attributed to anything else but the shape of a wave. This is what Helmholtz (1863/1870/1896) elaborated in great detail consequent to the basic statement that ‘sound colour’ (*Klangfarbe*) depends on the microstructure of vibration of a sounding body. This microstructure of course is reflected in the shape of each period of a sound wave recorded from an instrument. In regard to perceiving ‘sound colour’, Helmholtz (1863) argued that the ear is capable to perform decomposition of any complex periodic sound wave into its constituents according to Ohm’s law, that is, into sinusoidals (each of a given frequency and a given amplitude). Of course, Helmholtz knew the theorem of Joseph Fourier, according to which (in a brief interpretation Helmholtz gave as part of a popular lecture in Bonn 1857) any arbitrary [periodic] waveshape can be synthesized from a number of simple waves of different wavelength.⁷ Accordingly, Helmholtz held that our ear does the same what the mathematician does by applying Fourier’s theorem, namely “it dissolves [periodic] complex waves into a sum of simple (or elementary) waves”. Helmholtz (1870/1896) also explored in an approach one may call ‘analysis-by-synthesis’ production of complex periodic sounds (and even inharmonic sounds) from superposition of sinusoidals. The idea was to compare natural with synthesized sounds; if a synthesized sound came close enough to the original, its spectral composition could be stated in terms of the formula used in the additive synthesis.

⁷ Translations are from the edition of Helmholtz’ collected lectures and speeches (Helmholtz 1896).

3.2 A Matter of *Ausdehnung*: Stumpf on ‘Sound Colour’ (Klangfarbe)

The study of ‘sound colour’ was begun, as far as instrumentation and method is concerned, on a mechanical basis, namely by use of tuning forks and other devices as resonators, and of interference tubes for cancellation of partials. Along with analysis, empirical work on ‘analysis-by-synthesis’ was undertaken in that complex sounds were synthesized from a set of (almost pure) tones. Though scholars recognized that sounds can undergo marked changes not only in regard to pitch and dynamic level over time but also with respect to timbral characteristics, ‘sound colour’ was first viewed in a more static sense, namely as an attribute of sound that is present and audible for a certain time. As Stumpf (1890, 520ff.) has discussed in a phenomenological approach to perception of specifics of ‘sound colour’ (*Klangfarbe im engeren Sinne*), we can assign three basic attributes to sounds: height (*Höhe*), intensity (*Stärke*), and extension (*Größe*). While height and intensity can be directly related to physical parameters (frequency, amplitude), this is not the case for the third attribute that, in certain ways, for Stumpf has to do with perception of space and of spatial properties (cf. Stumpf 1890, 50ff.). The key word he uses in this context is *Ausdehnung* (that can be translated, in this context, as extension or volume), however, in his discourse there are several ‘spatial’ aspects to which he relates. These range from localizing the source of a sound in three-dimensional space, that is, a task performed in binaural hearing (a process that can be described in terms of acoustics and psychoacoustics), to a more subjective assessment of sound qualities such as ‘volume’, ‘density’, ‘brightness’, and ‘sharpness’. For example, musical tones low in fundamental frequency and ‘pitch’ seem to be more extended (to fill a larger ‘volume’, or to have a larger ‘body’) than high-pitched tones that, typically, appear as small and lacking ‘volume’. Further, low tones often appear as dull as well as soft while high-pitched tones appear as bright and also as sharp, etc. There are a number of such phenomenal attributes that, according to Stumpf (1890; also Stumpf 1926, Kap. 15) we use to characterize the quality of certain sounds we perceive. The attributes, however, are assigned to sounds in a quantitative way. If sounds are classified, for instance in terms of ‘thickness’, certain sounds may be rated as being ‘thicker’ than others.

Stumpf (1890, 535ff.) argued that ‘spatial’ attributes (such as massiveness or sharpness) apply even to pure (sine) tones, that those attributes are immanent to a tone of given height and intensity, and that these attributes vary in parallel with the ‘height’ and the brightness of tones.⁸ Also, certain combinations of the primary (objectively measurable) attributes tone height and intensity will result in certain

⁸ For an in-depth discussion of the tonal and sound attributes that were used by Stumpf, Hornbostel, Weltek, etc., see Albersheim (1939) and Schneider (1997b), Kap. III.1, 404–430.

sensations that are the basis of the ‘tone colour’ we assign even to pure tones.⁹ For example, a sine tone of high frequency and high intensity (SPL) made audible appears as sharp, or even as “piercing”. In general, sensation of pitch and brightness can be said to vary with the physical property of frequency (intensity is held constant), sensation of loudness varies with physical intensity (frequency held constant); density and sharpness vary (in upward direction) with frequency and intensity, volume varies against frequency, brightness and density (with maximal volume for tones low in frequency, ‘tone height’, and brightness). In regard to the issue: what is the “dimensionality” of sounds, it seems obvious that even pure tones have a number of physical properties and sensational as well as perceptual attributes that can vary in a quantitative way. For the physical properties (frequency, intensity [dB]), which can be varied along a continuum, and independent of each other, the notion of a ‘dimension’ applies. With respect to sensation and perception, the matter is much more difficult since, as is well-known from sensation of ‘pitch’ and brightness of pure tones, these are ‘integral variables’ whereas volume and brightness are (in principle) separable (cf. Schneider 1997b, 429f.). If the notion of ‘dimension’ is taken in a strict sense (requiring quasi-continua for measurement on at least interval scale level and correlational independence of variables so that such ‘dimensions’ representing variables can be unfolded as an orthogonal structure in a k -dimensional vector space), it will be hard to separate phenomenal attributes according to well-defined ‘dimensions’. Stumpf saw the interrelation of tonal attributes already in simple tones, for which ‘tone color’, he said (1890, 540) is not an attribute besides ‘height’ and ‘intensity’ but comprises “partly intensity, partly height, and partly extension” (or volume).

In regard to ‘sound colour’ (of complex sounds), Stumpf distinguishes characteristics he calls ‘inner’ moments from such he refers to as ‘outer’. The ‘inner’ moments relate to the structure of partials in a given sound, that is, the number of partials present and the relative amplitude of each partial. Hence, the ‘inner’ structure of sounds concerns the energy distribution in a (typically harmonic or, in certain types of sounds, also an inharmonic) spectrum. Inner moments of sound colour can be represented by a line spectrum as well as by a spectral envelope. The ‘outer’ moments relate to, first of all, temporal and dynamic aspects such as the onset and the decay of sound, the occurrence of transients and noisy components at the onset (as well as noisy components even in the steady-state portion), modulation and other fluctuations. Most of these phenomena can be described in terms of the temporal envelope. Taken together, ‘inner’ and ‘outer’ moments of ‘sound colour’ thereby form a three-dimensional structure (Fig. 4), in which x represents the partials (ordered by their harmonic number and/or their frequency [Hz or kHz]),

⁹ The term ‘tone colour’ (*Tonfarbe*, Stumpf 1890, 1926) applies to pure tones, while ‘sound colour’ (*Klangfarbe*) is reserved to complex tones. More empirical evidence for the fact that, by varying two physical properties (frequency, amplitude) of pure tones, one can vary more than two sensational attributes (namely, pitch, loudness, and volume), was later given by Stevens (1934).

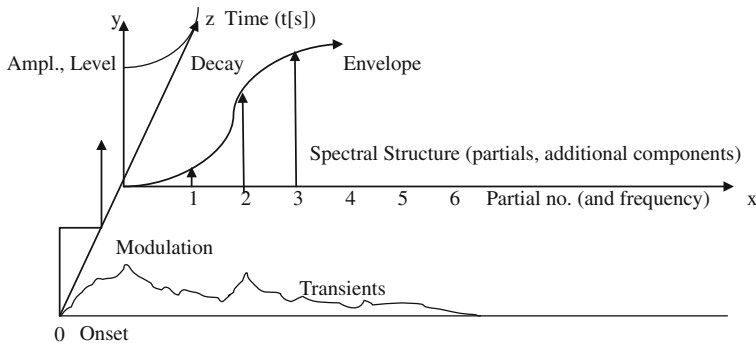


Fig. 4 ‘Inner’ and ‘outer’ moments of *Klangfarbe* (Stumpf 1926, Kap. 15)

y represents the (linear) amplitude of sound pressure or the intensity of the sound,¹⁰ and z represents time [s or ms].

Stumpf rightly emphasized that sounds produced from instruments quite often vary more in regard to the register in which they are located (from low or very low to medium to high or even very high depending on the ambitus or range of octaves an instrument can cover) than from one instrument to the other (the phenomenal difference for certain sounds produced by either a French Horn or a Cello in a low register thereby appears smaller than for two sounds produced by each instrument in either a very low or a very high register). This fact has led to considerations according to which one would need to keep the number and relative intensity of partials constant in order to maintain a certain ‘sound colour’ over several registers (cf. Slawson 1985). This idea (which owes to Slawson’s concept of composing sequences of ‘sound colour’ equivalent to sequences of tones in a melody) would imply shifting an identical spectral envelope along the frequency axis; such an operation would, however, not preserve zones of spectral energy concentrations that are often viewed as ‘formants’. Consequently, such a linear shift can have unwanted effects because sounds appear unnatural in timbre when envelopes are transposed by an octave up or down (cf. Rodet and Schwarz 2007, 177).

Another finding Stumpf reported was that even experienced musicians and instrument makers failed significantly to identify sounds produced from various instruments (presented at random) correctly if the onset including transients and the final decay were cut off from the sound, and only the quasi-stationary portion was audible for 2 s for the subjects. To be sure, such experiments again were carried out basically on a mechanical level, in this case, with sounds presented through tubes in a wall that could be rapidly opened and closed to cut out portions of sound (cf. Stumpf 1926, 374f.).

¹⁰ Where, approximately, $I = p_{\text{eff}} v_{\text{eff}} = p^2/Z$ that can be transferred to SPL [dB].

3.3 The Quest for ‘Formants’ in Musical Instruments

Helmholtz (1863/1870) devoted one chapter of his comprehensive work to differences various musical instruments show in regard to their respective *Klangfarbe*. He included a section on vowels as occurring in speech and in singing and reported many of his own observations plus some of the research done by other scholars. After going into details of resonance phenomena in the vocal tract and in particular in the mouth cavity, Helmholtz (1870, 179) came to the conclusion that vocal sounds differ significantly from most other musical instruments in that the relative power of partials is not dependent on their harmonic number but on their absolute pitch (or frequency) position.¹¹ As an example, he said that the vowel /A/, when sung on the musical note Es (E^b_2), would have a resonance peak at b'' (B_5), which is the twelfth partial of Es (E^b_2). If, however, the same vowel /A/ is sung on the note b' (B_4), there is still a peak at b'' (B_5) though in this case it is the second partial. In this respect, vowels apparently did differ from a range of musical sounds where the strength of partials typically decreases with harmonic number (such as $A = 1/n$ or in any other suitable ratio of amplitude to harmonic number).

The finding of Helmholtz, in a generalized form, implies that vowels retain spectral energy concentrations at certain frequencies or small frequency bands irrespective of the absolute fundamental frequency where phonation takes place. Though direct observation of phonation had become possible when Johann Nepomuk Czermak (1828–1873, a Czech-Austrian professor of physiology) had introduced the larynx mirror into the study of the vocal folds when in action, Helmholtz still saw the vocal tract (German: *Ansatzrohr*) consisting of the pharynx and the mouth cavity as the main part relevant for spectral shaping of vowels. Explaining vowel production basically by resonances in the vocal tract, Helmholtz (1870, 178) referred to the work of Robert Willis (1830), a professor of mechanics at Cambridge who had conducted relevant experiments with reed pipes of variable length and had claimed that each vowel can be related to a single resonance. Helmholtz's concept of vowels characterized by distinct resonances of the vocal tract and mouth cavity corresponding to certain musical pitches (e.g., /U/ to f [F_3], /A/ to b'' [B_5], /E/ to b'' [B_6]) was confirmed, in the main, by observations and measurements made by several scholars independently. It is not possible here to review the many notable contributions to the theory of vowels put forward, for the most part, by professors of anatomy and physiology like the Czech-Austrian Johann N. Czermak (1828–1873), the Dutch Franciscus Cornelis Donders (1818–1889), and the Germans Georg Meissner (1848–1905) and Ludimar Hermann (1838–1914).¹² Also of importance was the Finish-Swedish linguist Hugo Pipping (1864–1944) who published several relevant articles as well as a

¹¹ Helmholtz (1870, 179): ...*dass die Stärke ihrer Obertöne nicht von der Ordnungszahl derselben, sondern von deren absoluter Tonhöhe abhängt.*

¹² Meissner was a professor in Göttingen when conducting his research on vowels and sounds from musical instruments in the 1880s and 1890s. Hermann was a professor in the University of

monograph (Pipping 1894) on this matter and explicitly dealt with the sound colour of vowels as sung.

It should be noted that there were some controversies concerning the nature of the *formant*, a term coined by Hermann (1894, 267) who defined it as the “characteristic tone” within a spectrum. For many sounds, Hermann had calculated the formant as a ‘center of gravity’ (he called it *Schwerpunktmethode*) by taking three adjacent partials with their amplitudes. Since determination of frequency and amplitudes for partials could only be done by approximation (given the mechanical tools available for analysis, see also Meissner 1907; Herrmann 1908), Hermann came to the conclusion that the formant must not always stand in a harmonic frequency ratio to the fundamental but could be inharmonic as well. This provoked severe criticism because the formant was primarily understood as a resonance phenomenon in a tube (the vocal tract) filled with air that undergoes longitudinal vibration only.

From his own measurements, in line with those of Pipping (1890), Hermann concluded *dass die Höhe der hervorragenden Partialtöne der Vocale sich mit der Notenhöhe nicht wesentlich ändert* (Hermann 1891, 181). Since the formant should be kept fixed in a certain frequency band while the musical notes of a scale rise in fundamental frequency, Hermann (1894, 268) stated as a *Grundgesetz* (fundamental law) that *der Formant mit steigender Stimmnote in der Ordnungszahl immer weiter herabgeht, seine absolute Lage dagegen behält*.

Following to the investigations of Helmholtz in regard to the sound colour of musical instruments and the human voice, it was proposed that the basic sound quality of certain instruments could be compared to vowels, whereby the bassoon should be similar to /U/, the French horn to /O/, the trombone to /A/, the oboe to the German /Ä/, and so on (cf. von Qvanten 1875). In much of the early research, the formant was viewed either as a single harmonic partial of strong intensity (the most prominent resonance in the vocal tract or in a tube filled with air such as a flute or flue pipe) or as kind of a *Mundton* (as Hermann claimed to exist) resulting from a separate regime of vibration. However, it had become clear quite soon that the specific quality of vowels as well as of sounds characteristic for a certain musical instrument resulted from groups of partials rather than from single spectral components. Meissner (1907, 595), summing up his findings, stated that (1) it is groups of higher partials with significant amplitudes that characterize both the sound of wind instruments (aerophones, *Blasinstrumente*) as well as vowel sounds of the human voice. He had found (2) such concentrations of spectral energy not to shift with different musical notes played or sung (that is, such concentrations are relatively independent of the ‘pitch’ played). Meissner concluded that the groups of relevant partials (3) are forming regions of spectral energy concentration which are fixed in frequency (*Eine den Klang eines Blasinstruments wesentlich*

(Footnote 12 continued)

Königsberg and internationally acknowledged as author/editor of textbooks on physiology. Some of the historical background to this era of research is briefly summed up in Graf (1980, 211ff).

charakterisierende Gruppe höherer Obertöne ...ist ein festes Gebiet oder eine feste Region bestimmter absoluter Tonhöhen...). In research done in the 20th century, the term for the regions Meissner had found usually was *Formantregionen* or *Formantstrecken* (cf., e.g. Stumpf 1926; Vierling 1936; Winkel 1960).

One interesting point reported in several of the early publications (e.g., Meissner 1907; Herrmann 1908) is that, in vocal sounds such as vowels as well as in sounds recorded from various pipes, the fundamental f_1 was found quite weak or almost missing. This led to questioning the Ohm/Helmholtz-theory of pitch as applicable to all kinds of sounds and to suggesting that ‘periodicity pitch’ as proposed by Seebeck might be a valid alternative in certain cases. Further, it was found that air in a cylindrical tube excited by a reed (as a valve periodically opened and shut) did not yield a sound spectrum confined to odd harmonics since even harmonics with significant amplitudes were measured as well.

When Stumpf finally wrote his book on speech sounds (*Sprachlaute* such as vowels and consonants, Stumpf 1926), he could draw on a broad range of previous research plus a wealth of observations and data he and his co-workers had collected. Stumpf offered a descriptive and analytical treatment of many phenomena relevant for phonetics though he (1926, VI) considered this book mainly as a continuation of the chapters on *Klangfarbe* he had offered in vol. II of the *Tonpsychologie* (1890, § 28). In his *Sprachlaute* monograph, Stumpf also included chapters on the psychology of listening and on topics relevant for psychoacoustics. Concluding his book, he added a chapter on instrumental sounds (Kap. 15: *Über Instrumentalklänge*) in which he treated the problems of ‘sound colour’ once more (on the basis of experiments he had conducted for many years after the second volume of the *Tonpsychologie* had been published).

Naturally, one of the chapters on the analysis of sung vowels relates to ‘formants’ (Kap. 2, 62ff.). Formants according to Stumpf are not necessarily confined to single partials (as most of previous research from Helmholtz to Hermann had suggested) but rather may include several partials so that energy is concentrated in a frequency band (*eine Strecke des Tongebietes*). Stumpf distinguished between a *Hauptformant* (the frequency band in which the most prominent spectral peak is found) and one or several *Nebenformanten* that appear as additional relative maxima of spectral energy distribution and can be found above or below the *Hauptformant*. As to the debate whether formants are rising in parallel with f_1 of the notes sung or played on instruments (usually within one octave) or are almost fixed in their frequency position, Stumpf (1926, 62ff., 191f., 377f.) did not accept the view of Hermann (1891) concerning “frequency-fixed formants” but took kind of an intermediate approach in that he argued that formants are only relatively stable in their frequency position, and will shift in the direction of the notes played or sung in a scale: *ihre Bewegung erfolgt in gleicher Richtung, aber weit langsamer, sie umfaßt (abgesehen vom Grundton c_2) nur wenige Töne*. Thus formants should shift only slightly with rising ‘pitch’, and only within certain boundaries. Probably due to his focus on vowels, Stumpf reported spectral energy maxima for a number of instruments (in particular woodwinds and brass) which he, in analogy to speech, addressed as formants. He saw a main formant (*Hauptformant*), typically

comprising harmonic partials no. 4–6, shifting relative to the fundamental frequency of different musical tones that were played, while several *Nebenformanten* (adjunct formants) that might also occur according to his observations should remain within a certain frequency range.

It should be noted that, after Hermann's *Schwerpunktmethode* for finding the center of formant regions, a new method was proposed by Vierling (1936) on the basis of the phase relationship between several partials making up a formant group for which a spectral envelope with a single peak can be found in an amplitude spectrum. Taking the frequency corresponding to an envelope maximum as the resonance frequency ω_{res} (for a resonance filter in a generator/resonator model), its phase would be zero; it follows from an equivalent resonance filter curve that for all partials below ω_{res} (as the point of phase inversion), the phase is negative, and for all partials higher in frequency, the phase is positive. Vierling suggested that, by superimposing the partials with correct amplitudes and phases, a new wave results, whose envelope periodicity frequency would represent the center of the formant region.

Not long after *Die Sprachlaute* had been published, Stumpf's former student Karl Erich Schumann submitted his Habilitationsschrift on *Die Physik der Klangfarben* (typewritten, 1929) to the Berlin university.¹³ Schumann, who—from what is known today as factual evidence—took his Ph.D. with Stumpf (in 1922),¹⁴ had specialized in musical acoustics and had published a small book on this matter in a popular scientific series. His treatment of formants in this publication (Schumann 1925, 76–79) sums up some of the earlier findings and discussions (Helmholtz, Hermann) and refers briefly to experiments Stumpf had conducted on the synthesis of vowels. It is not possible, at this place, to discuss Schumann's work in any detail though it became influential in some circles in the 1970s, and has been referred to more frequently since (see Mertens 1975; Fricke 1993; Reuter 1995; Reuter 1996). Ironically, Schumann's *Physik der Klangfarben* never got published due to, it seems, effects of World War II (while Schumann played an important role in the organization of the Nazi war effort and in the development of weapons, in particular high-grade explosives; cf. Nagel 2007).

Schumann had postulated four *Klangfarbengesetze* (laws reigning 'sound colour'; see Mertens 1975; Reuter 1996, 110ff.) which all relate to formants in the spectra of musical instrument sounds. Though there are certain patterns or even

¹³ A copy of this Habilitationsschrift is kept in the library of the Staatliche Institut für Musikforschung von Berlin. Another copy apparently is in the Institute of Musicology at Cologne.

¹⁴ As to Schumann's biography, see his Wikipedia entry and Nagel (2007, 232ff). Stumpf became an emeritus in 1921 but continued to teach (including supervision of candidates) until late in the 1920s. According to Reinecke (2003, 185) who knew Schumann personally, he was *Schüler und langjähriger Assistent Carl Stumpfs*. However, Stumpf (1926, 7), in the *Einleitung* to his *Sprachlaute*, mentions eleven co-workers and colleagues who had helped him in experiments on the analysis of vowels (notably Dr. von Allesch), but not Schumann who (cf. Nagel 2007, 232ff.) apparently was a paid assistant in the Institute of Physics where he worked with Arthur Wehnelt, well known for his discoveries in electronics.

regularities that can be found empirically in the structure of spectra of certain types of instruments (e.g., reed-driven aerophones, see Voigt 1975), the notion of a ‘law’ would require generalization that, from the evidence at hand, seems hard to justify yet. The concept of ‘formant’ itself has been given somewhat different interpretations in regard to vowels and musical sounds (see above). In a general view, *the term is synonymous with resonances in the vocal tract* (Slawson 1985, 38). According to acoustic fundamentals of speech production in a system that can be described as a source/filter model comprising a generator (the vocal folds) and a resonator (the vocal tract up to and including the lips), formants are regarded as resonances in a cylindrical tube of length L (cm) closed at one end (a $\lambda/4$ resonator). Hence, the resonance frequencies should be found at $F1 = c/4 l$, $F2 = 3c/4 l$, and $F3 = 5c/4 l$ of a resonator of length L with $c =$ speed of sound in air (see Neppert and Pétursson 1986) where $F1$, $F2$, $F3$ represent the three (most prominent) formants. For the sound pattern corresponding to the neutral vowel /ə/, and the typical male resonator tube of $L = 17.5$ cm length, the formant frequencies should be close to 500 Hz, 1500 Hz, and 2500 Hz, respectively. This is an idealization, though, since both the generator and the resonator can be varied dynamically by speakers or singers during phonation. As has been demonstrated by Fant (1960) by means of measurement and modeling, the cross section profile and even the length of the vocal tract (German: *Ansatzrohr*) as well as other parameters relevant for ‘pitch’ and spectral energy distribution can be modified in the course of phonation processes. There is a range of articulation effects resulting from small changes in the position of the larynx, the lower jaw, the tongue (body and tip as two separate subsystems), and the lips (for details, see Sundberg 1997) which lead to modifications also of formants in regard to their bandwidth and frequency position relative to a fundamental as well as to changes in spectral energy distribution and spectral envelope of sound radiated from the mouth.

The distribution of spectral energy for an individual singer and a particular vowel sung with the ‘colour’ of a certain language can be determined by analysis of steady-state portions of sounds. For example, taking the long note sung by a woman on the vowel /o/ at the beginning of her rendering *Abu Zeluf* (see the left half of the melogram in Fig. 2), one gets a fairly stable spectrum with a pattern of peaks shown in Fig. 5.

The sound actually contains spectral energy up to at least 12 kHz. The fundamental is found 333 Hz, with the second harmonic (667 Hz) strongest in amplitude. Taking the spectral envelope in Fig. 5 calculated from a formant filter analysis, one finds peaks at harmonics nos. 2, 4, 9 and a zone of condensed spectral energy carried with the harmonics 11, 12, 13. Taking a template for vowel formants (cf. Födermayr 1971, Abb. 21), one sees that for the vowel in question energy maxima should occur in four frequency bands: the first is around 700 Hz, the second around 1 kHz, the third covers a band from, roughly, 1.8 to 2.7 kHz, and the fourth a band from about 3.3 to 4 kHz.

As had been stressed already in earlier research, certain sounds produced from various musical instruments appear as similar to vowels when such are sung. Phenomenal similarity of course can be rated by subjects, and especially by those

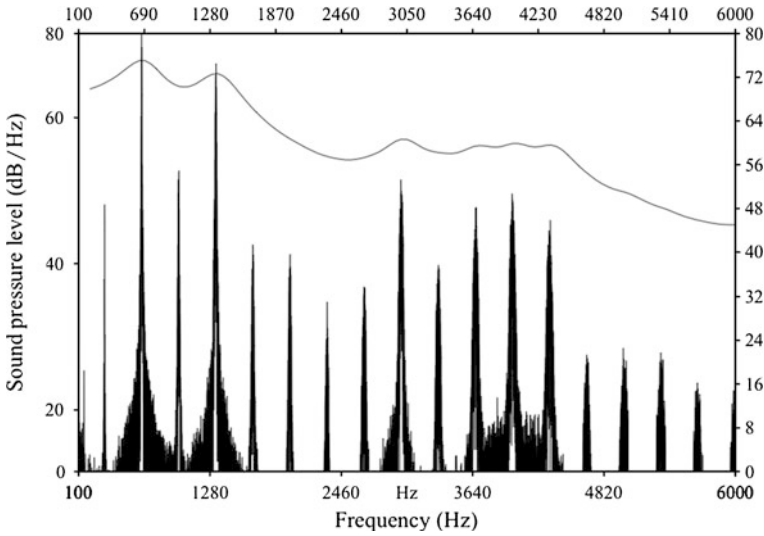


Fig. 5 Spectrum and formant filter envelope, vowel /ɒ/ as sung by Dunya Yunis

with a musical training. However, ‘vowel quality’ in violin sounds can be detected also with signal processing tools where the output is referenced to a Jones-type diagram of vowel categories (see Mores 2010). Since subjective ratings of the ‘vowel quality’ in violin sounds and automated extraction of vowels show a high degree of convergence, there must be distinctive features which permit to classify sounds as to their ‘vowel quality’. In retrospect, this has led to the issue of ‘formants’ as well as to several explanations why formants are genuine to musical sounds. Given that the steady-state of most vowels produced in singing shows a pattern of peaks and/or concentrations of energy (the latter may be connected with a single partial or with a group of partials) in the envelope, one can address formants accordingly as (1) single prominent spectral peaks, as (2) groups of partials with amplitudes or intensities above the average level of neighbouring partials, or (3) simply as frequency bands in which more spectral energy is found than in adjacent regions of the spectrum. The “ideal” formant spectrum, then, would be characterized by a regular pattern of spectral maxima and minima so that the envelope would show a cyclic structure defined by zeros. Such spectra can be obtained from trains of rectangular pulses with a duty cycle of τ/T (where τ is pulse width [ms], and T is the pulse period[ms]). Amplitudes of partials in the spectrum of a train of rectangular pulses conform to a sinc function ($\sin [x]/x$) where the zeros are found at $n \tau/T = 1, 2, 3, \dots$ (cf. Meyer and Guicking 1974, 40f.). The function useful for demonstrating a cyclic spectrum would be of the form $\text{Sin}[x]^2/x$ so that the envelope is all positive (Fig. 6); the roots of the function of course are found at $n \pi, n = 1, 2, 3, \dots$. The peaks and troughs of the envelope can be conceived as covered by another envelope that represents the overall exponential decay of amplitudes (dashed line):

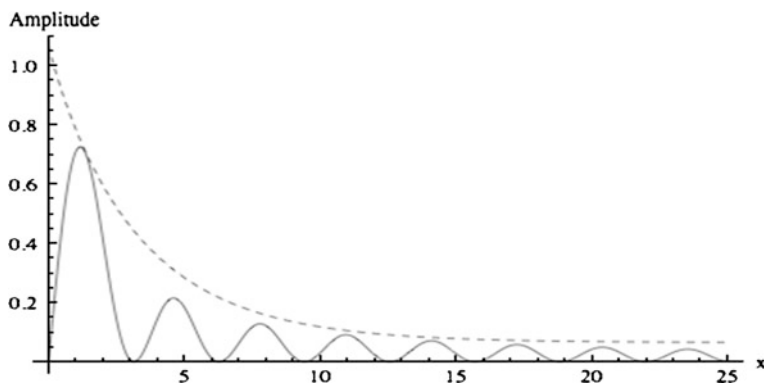


Fig. 6 Model of a cyclic spectrum (envelope of the form $\text{Sin}[x]^2/x$)

Since spectra obtained for sounds of certain musical instruments such as reed-driven aerophones basically exhibit a more or less cyclic structure [as is the case for a number of double reeds (cf. Voigt 1975) and also reed pipe stops in organs, see Beurmann et al. 1998], the explanation at hand is taking the reed as a pulse generator that should produce, in theory, a sequence of rectangular pulses where the pulse of height A and duration t corresponds to the time the valve is open so that a flow U of air at a certain pressure p is released into the resonator while the time between pulses marks the duration for which the valve is shut. Though such a model serves to explain the basic principle, things are quite complex in reality since even for instruments with a beating reed such as the clarinet the vibration of the reed and the pressure measured in the mouthpiece (cf. Backus 1963) do not yield a rectangular shape equivalent to a pulse (which is defined as a [discontinuous] jump-function). Rather, reed vibration and pressure seem to be almost sinusoidal when excitation of the reed is very soft, and approximate a triangular shape with medium blowing pressure. Only when driven hard, motion of the clarinet reed becomes more of a rectangular wave. It was observed that for soft blowing the valve never completely shuts while it shuts for about a half-cycle when high blowing pressure meaning a very strong force acting on the valve is applied.

If we take a comparatively large double reed mounted on a small brass tube as it is used for the French-Breton *bombarde*,¹⁵ the reed generator produces a periodic change in pressure at the output (see Schneider 1998, Figs. 1, 2) that becomes audible as the source signal since the double reed is fixed to a small cylindrical

¹⁵ The triangular double reed used for measurements is 10.5 mm wide at the opening. It is 20 mm long and is mounted on a cylindrical tube of 25 mm length and 4 mm diameter. Thereby, the total length of reed and tube is 45 mm. Recordings were made with the reed and tube put close in front to a condenser microphone (Neumann U 67, AKG C 414 B-TL II) set to cardioid and fed into a preamp and A/D converter system (Telefunken V 76, Panasonic DAT SV 3800 at 16 bit/48 kHz and RME Fireface 800 at 32 bit float/96 kHz on hard disc.

tube of 25 mm length that already may act as a (first) resonator. The *bombarde* can be modeled as a coupled system comprising the valve as such, this small brass tube, the conical bombarde pipe of 27.5 cm length (with 8 mm bore at the upper, 15 mm at the lower end) plus the bell which has an effective length of 3.4 cm and opens from 16 mm to 48 mm. For the double reed on the small brass tube, at soft to medium pressure force, the wave shape of the source signal is approximately triangular (comparable to the beating reed generator of the clarinet) while the signal becomes more complex with respect to its fine structure per period with increasing blowing pressure. The reed generator (double reed bound to its proper tube) for medium excitation yields a spectrum comprising a series of quasi-harmonics starting at ca. 698 Hz as listed in Table 1.

Table 1 *bombarde*, reed generator output spectrum 0.5–14 kHz (FFT: 16384 pts, Hanning)

Partial no.	Frequency	dB ^a	Partial no.	Frequency	dB
1	698	-59.4	11	7,658	-73.3
2	1,395	-13.4	12	8,372	-43.7
3	2,094	-54.8	13	9,063	-73.6
4	2,790	-27.4	14	9,767	-42.7
5	3,510	-77.9	15	10,446	-73.7
6	4,186	-17.8	16	11,162	-42.2
7	4,837	-66.0	17	11,853	-72.5
8	5,581	-41.3	18	12,556	-52.0
9	6,248	-74.2	19	13,247	-83.3
10	6,976	-34.4	20	13,957	-62.8

^a SPL relative to 0 dBFS; therefore, all dB readings for undistorted signals are negative (from 0 to -100 dB)

There are some more partials above 14 kHz up to 22 kHz so that excitation of modes in the resonator is secured even at modest blowing pressure. It is obvious that partials no. 2, 4, 6 are exceedingly stronger than their neighbours, partials 1, 3, 5 and 7, so that they will make up good candidates for ‘formants’. In fact, if the reed generator sound is put through a formant analysis using the routine implemented in the Praat software,¹⁶ it turns out three ‘formants’ (Fig. 7).

One can see that the three ‘formants’ in fact correspond to partials no. 2, 4 and 6 of the reed generator spectrum (Table 1), and that there is considerable energy fluctuation in partials 2 and 6 over time while partial 4 remains stable.¹⁷ Whether one may call the three strongest partials of the generator spectrum ‘formants’ because they ‘stand out’ against their neighbours seems a matter of terminology

¹⁶ The analysis is based on STFT with an approximation of spectral peaks in the envelope per analysis frame by the Burg method (cf. Marple 1987). The three ‘formants’ in fact represent tracks of spectral envelope peaks plotted against time.

¹⁷ The fluctuation is evident from so-called speckles (black dots) which indicate spectral energy peaks per analysis frame.

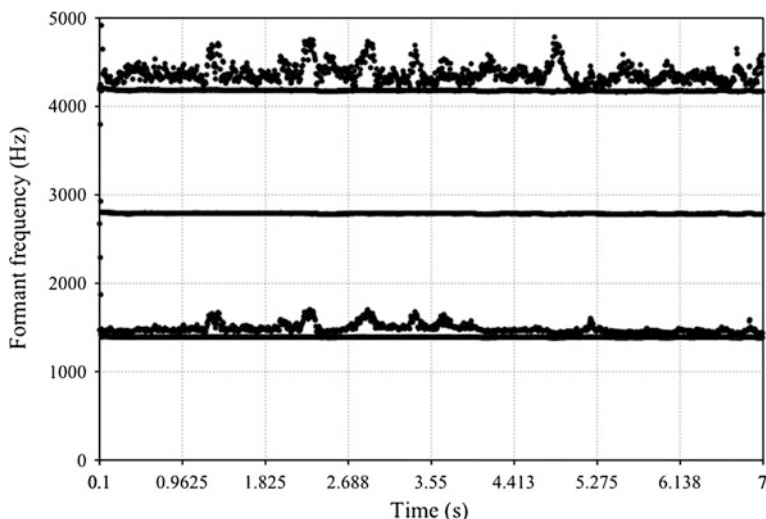


Fig. 7 *bombarde* reed generator output, formant analysis (Burg algorithm)

rather than of principle. However, in this case each of the three ‘formants’ of the reed generator would consist of but one partial.

In case the generator is excited with maximum admissible blowing pressure (there is of course a limit beyond which the two reeds are pressed against each other and the valve simply remains shut), crests of the wave become quite steep and the wave shape even more complex per period due to additional modes of vibration.¹⁸ Consequent to overblowing, the bandwidth of the source spectrum in this condition starts with partial no. 2 indicated in Table 1, and increases with blowing force and pressure applied to the valve so that the source spectrum of the reed generator carries energy up to about 33 kHz. The spectrum shows distinct and strong peaks up to 10 kHz from where on peaks are broader; also, there is considerable energy distributed in between peaks due to noise from the air flow through the valve. The reed generator spectrum however does not reveal a cyclic structure with clear dips or gaps at certain partials as one would expect from a system that produces rectangular pulse trains as output.

If driven with a medium blowing force applied to the reed generator, *bombarde* sounds in the normal playing range (i.e., not overblown into the higher octave) have spectra like that displayed in Fig. 8:

¹⁸ In reed-driven instruments such as the clarinet, a range of nonlinearities is observed with respect to the vibration of the (single or double) reed, the flow of air through the slit as well as in other parameters. For details in regard to modeling and calculation, see Dalmont et al. (2000), Kergomard et al. (2000).

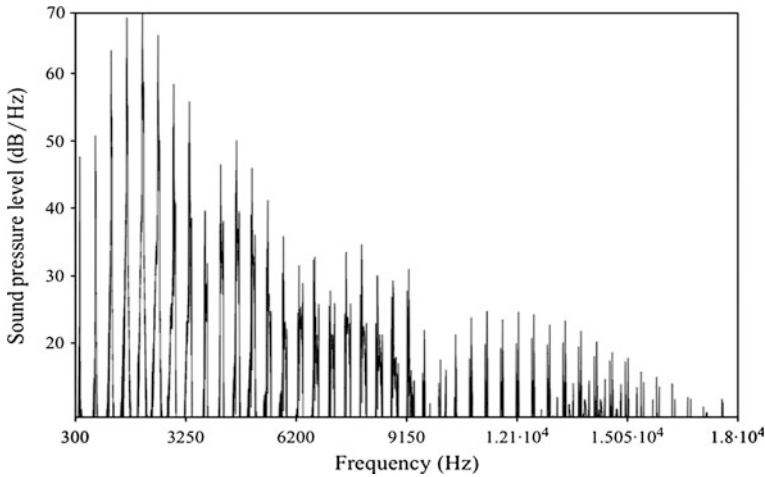


Fig. 8 Spectrum, *bombarde* note 1, $f_1 \approx 419$ Hz (0.3–18 kHz, 42 partials displayed)

The spectrum can be structured into four main groups of partials, which cover (I) partials 1–9, (II) 9–15, (III) 15–24, and (IV) 24–41 so that the partials two groups have in common are spectral dips that separate two neighbouring groups from each other. If an envelope is put on the partial amplitudes for each group, it will approximate somehow a shape like \cap (an inverse U), however, the spectrum hardly can be called cyclic in any strict form (according to the model in Fig. 6). Though sound spectra for most notes of the *bombarde* played within one octave are similar in structure, there is still some variation which is also evident from spectral statistics given in Table 2 (all sounds normalized to -3dB before analysis):

Table 2 *bombarde*: notes 1–8, f_1 , spectral center of gravity and SD^a

Note/tone	f_1 (Hz)	Center of gravity(Hz)	SD(Hz)
1	419	2082.77	642.83
2	452	2156.88	528.73
3	502	2098.26	804.65
4	566	2451.21	847.87
5	597	2561.56	1101.84
6	673	2720.42	1051.01
7	752	2824.61	1317.79
8	839	2928.04	1225.11

^a The center of gravity is the average of f over the frequency band covered by the analysis, weighted (in this case) by the power spectrum (for details, see documentation in the Praat manual)

The data with the center of gravity rising from one note and sound to the next (except for note/sound 3) indicate spectral energy basically seems to shift with the rise of f_1 rather than to remain fixed in a certain frequency position. If segments of

one second each of the sounds 1–8 are subjected to a formant analysis (Burg method), there will be up to five, and typically four, tracks that are created in the frequency range relevant for formants with respect to a male (4.5 kHz) or a female voice (5.5 kHz) (Fig. 9).

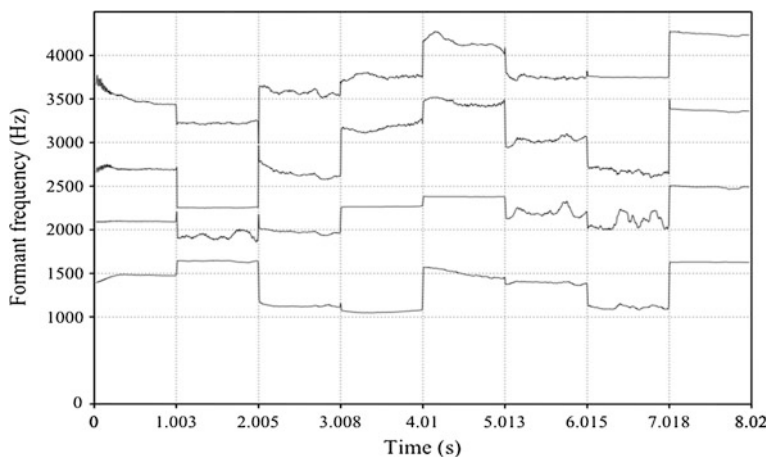


Fig. 9 Formant analysis (Burg), segments of *bombarde* sounds 1–8

One can mostly relate the four formants per sound to particular partials or to pairs of such partials. For instance, in the first sound segment the lowest formant at 1.5 kHz is very close to the strong partial no. 4 of the spectrum (Fig. 8) at 1674 Hz/68.2 dB, formant 2 represents the most prominent partial of the spectrum (no. 5 at ca. 2093 Hz/68.7 dB), formant no. 3 is in between partial no. 6 (ca. 2511 Hz/65.6 dB) and no. 7 (2929 Hz/58.6 dB), etc. Whether there are any “formant laws” inherent in the sound especially of reed instruments (cf. Mertens 1975; Voigt 1975) that could be derived from spectral analysis at present seems hard to tell; in any event, one would have to analyze a sample of many sounds to justify such a conclusion in regard to principles of inductive generalization.

Though reasonable approximations to cyclic spectra can be found in the sound of reed-driven instruments (and in particular in organ reed pipes, cf. Beurmann et al. 1998; Schneider et al. 2001), there are probably much closer approximations to cyclic spectra at hand for plucked strings of harpsichords where the harmonics that are cancelled out are determined by the ratio of the string length and the plucking point, L/l (for examples, see Beurmann and Schneider 2008, 2009). An approximate cyclic spectral envelope of course indicates that there are bands where energy is concentrated relative to the dips or gaps in between. Following common terminology, one may address such concentrations of energy in groups of partials as ‘formants’ even for harpsichord sounds. It would need empirical evaluation involving experienced listeners to find out to which extent such sounds in perception might appear as similar to sung vowels. After all, it should be remembered that the concept of ‘formant’ was developed for *vowels* as observed in *singing*, in the first place.

3.4 *Stumpf Reconfirmed: Periodicity, Harmonicity, Verschmelzung, Consonance*

‘Inner’ and ‘outer’ features of *Klangfarbe* condensed into Fig. 4 are suited to characterize a large number of sounds in an objective way (namely, by signal analysis as well as through psychoacoustic experiments). In regard to objective criteria, Stumpf (1926, 390) identified *Klangfarbe im engeren Sinne* with spectral structure, and subjectively with the sum of so-called *Komplexeigenschaften* resulting from listening to sounds composed of partials. To Stumpf (1926, 278), sounds comprising a number of partials (and in particular vowels) in normal perception (i.e., holistic perception not directed to structural analysis) will appear as complexes that are not segregated or ordered (*ungegliederte Komplexe*). To Stumpf, sound colours were classic examples of *Komplexeigenschaften* since they are perceived as wholes that must not necessarily be analyzed into their constituents. A different matter though is a complex of partials that is experienced as giving a special perceptual quality, that of *Verschmelzung* into a highly consonant formation (cf. Stumpf 1890, 1898, 1926). Stumpf himself had experienced highly consonant complexes of partials by listening to organ mixture stops as well as by synthesizing vowels; applying attentive listening to such sounds, he regarded *Verschmelzung* an experience that reflects not just “fusion” of partials but even more so their interpenetration (*Durchdringung*; cf. Stumpf 1890, 128ff.). Stumpf’s view was that the perception of spectral *Verschmelzung* and *Durchdringung* must have a neural basis, and that apperception of consonance would most likely be effected on the cortical level.

It is characteristic of such complexes (cf. Schneider 1997a, b, 2000) that perception can switch between two different modes, one being holistic and one analytic.¹⁹ While the holistic experience of such a complex of harmonic partials as produced by a chord of three or more complex tones in just intonation (see examples in Schneider 1997a; Schneider and Frieler 2009) rests on facts that can be described in terms of acoustics and psychoacoustics (namely, strict periodicity of the time signal, clear microstructure of the waveshape with steep crests at the beginning of each period, absence of beats and roughness, strict harmonicity of the spectrum, strong difference and combination tones present at appropriate SPL), the analytic approach can be directed to segregating the complex sound into several constituents whose relational structure is perceived and cognitively evaluated (this is one of the reasons why Stumpf and other psychologists with a philosophical background maintained to use the notion of apperception for such a process; cf. Stumpf 1926, 279, 372; Schneider 1997a, b). To illustrate the case, we synthesized a sound Stumpf himself (1926, 394) has proposed for demonstrating the effect of a special concord (*Zusammenklang*) that is perceived as *sehr einheitlich, aber noch reicher und von wunderbarer Schönheit* (even more coherent and richer in

¹⁹ The difference is not identical with, yet bears some parallels to, that between bottom-up and top-down analysis as outlined in Bregman’s *Auditory Scene Analysis* (Bregman 1990).

harmony than another concord that had been proposed by Dayton Miller). Stumpf's concord comprises five perfect triads played simultaneously that are built on the root frequencies of 100, 200, 400, 800 and 1600 Hz, respectively. Hence, we need 15 partials, which are arranged so as to match Stumpf's desired spectral structure (that included an envelope where the intensity decreases regularly from one partial to the next; in our sound, damping per partial is set to -3dB). Also, when synthesizing the sound (done with Mathematica) we put a temporal envelope on the partials for a smooth decay just as Stumpf had suggested (Fig. 10).

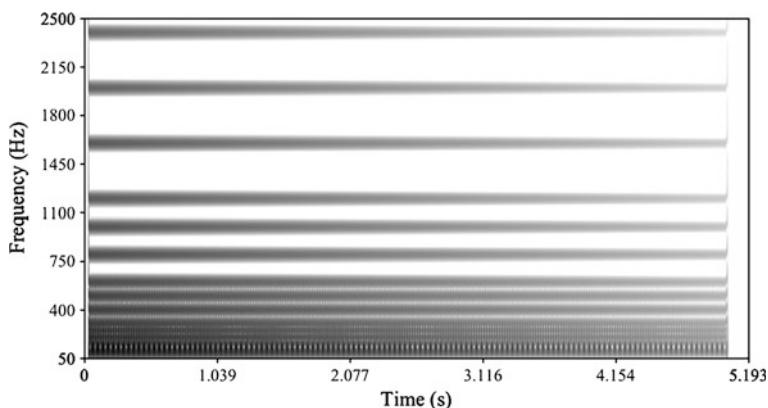


Fig. 10 Stumpf's perfect *Zusammenklang*, 5×3 partials = 5 just triads

Hearing such a complex, the attentive listener indeed can switch between a holistic mode where he or she will perceive a chord-like sonority of fifteen tones, and a more analytic mode directed to evaluating the spectral structure of the complex as well as the tonal relations between the five triads. Though there are certain limits for such an analysis as far as frequency resolution of the ear and, consequently, identification of components by means of perception is concerned, the pitch structure and also the interplay of some of the partials in this complex can be recognized. The basis for a cognitive evaluation is an analysis that takes place in the auditory pathway consequent to two interacting mechanisms, peripheral spectral analysis and the detection of periodicities arising from spectral components as well as from the temporal envelope. The autocorrelation process (ACF) will also reveal periodicities corresponding to virtual pitches including a series of subharmonics.²⁰ Hence when fed into a model of the auditory periphery working in the time domain that processes complex sounds from basilar membrane filtering and hair cell transduction to neural nerve spike representation in the auditory nerve

²⁰ Not to be confused with the subharmonic matching process that has been proposed by Terhardt (1998) for 'pitch' (f_0) estimation of complex sounds; of course, there are some correspondences in both concepts.

(see Meddis and O'Mard 1997), the output aggregates periodicities found in the neural activity pattern (NAP) into a sum ACF (SACF). As can be expected, the neural output contains all periodicities inherent in the sound input plus virtual pitches and subharmonics (see Schneider and Frieler 2009). Exactly this will happen also with Stumpf's 'paradigm sound' (1926, 394: *idealer Klang*) since an analysis performed with various algorithms including standard and advanced autocorrelation as well as cepstrum analysis (see Mertins 1996, 1999; Arfib et al. 2002) not only turns out lags (ms) corresponding to actual partial frequencies (such as 10 ms = 100 Hz) but also gives the two main periods (10, 25 ms) governing this complex sound. In addition, the analysis yields a series of subharmonics below the fundamental ranging from 50 Hz down to 10 Hz. Hence we confirm once more (cf. Schneider 1997a; Schneider and Frieler 2009) what Stumpf had experienced and what could be expected taking fundamentals of acoustics and psychoacoustics into account: sounds composed of harmonic partials organized into several complex tones representing a concord in just intonation will result in the perception of a highly consonant formation having a distinct 'Gestalt' and sensory quality. The explanation on the level of the sound signal of course is the causal relation between strict periodicity of the time signal and perfect harmonicity of the spectrum as defined by the Wiener-Khintchine theorem (cf. Meyer and Guicking 1974, 110ff.; Hartmann 1998, Chap. 14).

The following example should demonstrate the robustness of the principle: consider three cosine pulse trains which have fundamental frequencies at 300, 400, 500 Hz plus a number of harmonics so that equal energy is contained in the spectrum at {300, 400, 500, 600, 800, 900, 1,000, 1,200, 1,600 Hz}. The pulse trains (sampled at 16 bit/44.1 kHz) are in a harmonic ratio and might be regarded as representing both a sound signal and the corresponding neural output. The time function $s(t)$ arising from the superposition of the pulse trains for 100 ms is shown in Fig. 11 (for an almost identical SACF output derived from processing in an auditory model, see Schneider and Frieler 2009, Fig. 3).

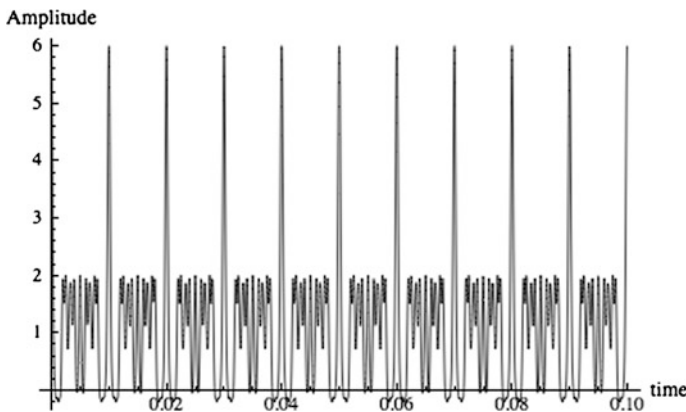


Fig. 11 Periodicities inherent in pulse trains based on harmonic ratios 3:4:5

From Fig. 11 it is easy to see that the signal is strictly periodic with a period $T = 10$ ms implying a repetition frequency f_0 of the compound pulse of 100 Hz that constitutes the ‘root’ of the harmonic series and gives rise to the sensation of a salient low virtual pitch. The complex sound thus composed of course is highly consonant in regard to Stumpf’s concept of *Verschmelzung*.

If the pulse trains are subjected to frequency modulation (FM) with individual modulation frequencies per Fourier component, the resulting sound is audibly shifting up and down in pitch and spectrum so that the degree of *Verschmelzung* one may assign to the percept is much less than that of the unmodulated sound (Fig. 11), and the time function $s(t)_{\text{fm}}$ of the compound pulse at a first glance appears quite irregular (Fig. 12; 100 ms displayed).

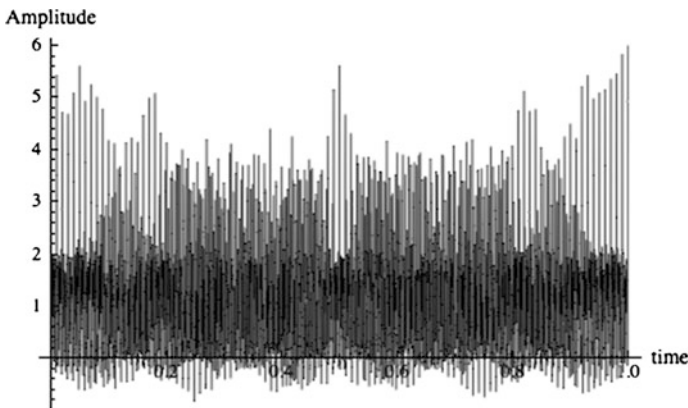


Fig. 12 Pulse trains with fundamental frequencies in the ratio of 3:4:5 subjected to FM; all Fourier components have individual modulation frequencies $A_n \sin(2\pi kt)$, $k = 1, 2, 3, \dots$

Though it is not easy to detect the periodicities still inherent in the modulated time function $s(t)_{\text{fm}}$ from looking into the graph, autocorrelation analysis still finds some of the Fourier components making up the compound as well as f_0 clearly marked at 100 Hz. The reason is that FM in this sound also is periodic (though with different modulation frequencies applied to different components), and that the autocorrelation function (ACF) by Norbert Wiener (1961) explicitly had been defined as the temporal mean of the product $s(t) s(t + \tau)$ in order to detect periodicities even in complex and/or noisy signals (in particular, EEG recordings). Hence, ACF and, similarly, cross-correlation (CCF; see Hartmann 1998, 346ff.; Ingle and Proakis 2000) such as used to compare two sequences (e.g., two audio signals or one audio signal recorded from both ears) is a tool suited to detect periodicities in various signals; ACF and CCF is robust in regard to angular modulation (frequency, phase) as long as the modulation itself is periodic (or nearly so). Therefore, pitch extraction based on ACF or CCF works also for slightly detuned intervals (as is the case in equal temperament) or even for inharmonic signals up to a certain degree of inharmonicity (since increasing inharmonicity of the spectrum implies decreasing periodicity of the time signal).

As far as auditory perception is concerned, Licklider (1951, 1956) it seems was the first to draw on ACF for a model of pitch perception where *the autocorrelation mechanism consists of many delay-line autocorrelators in parallel*. In addition, Licklider proposed a cross-correlational operation conceived as a *time-coincidence arrangement with two inputs and two delay lines running in opposite direction*. Though the basic idea of time-domain pitch analysis (that had been proposed by scholars such as Seebeck and Schouten) has been widely accepted and has been realized in many models of “neurally inspired” auditory perception (for an overview, see de Cheveigné 2005), the ACF-based model suffered from a lack of convincing neuroanatomical and neurophysiological evidence since an array of neural delay-line autocorrelators was not yet discovered. Also, processing speed in the auditory pathway is at odds with the huge delay needed to perform ACF for low pitches. On the other hand, in neurophysiological research undertaken from the 1960s onwards (summarized in Hesse 1972; Keidel 1975; Ehret 1997; Schneider 1997a, b; Nelken 2002), there had been measurements of periodicities in the auditory nerve as well as on the level of higher relays of the auditory pathway (notably, the ICC and the CGM) corresponding to periodicities in input signals. More recently, the neural basis for differences between consonant and dissonant musical intervals has been demonstrated by all-order interspike interval histograms recorded from auditory nerve fibers of cats in animal experiments (Tramo et al. 2001; see also Cariani 2004). Also, an improved auditory processing model that includes the auditory nerve, the cochlear nucleus (CN), and the inferior colliculus (ICC) has replaced the ACF process by operations on a huge array of components which are *physiologically plausible* (Meddis and O’Mard 2006).

3.5 Transients and Dynamic Evolution of Sound Spectra

In addition to finding pitch curves based on either f_1 or f_0 measurements (see above), spectral structure and spectral energy distribution have always been of interest. After decades of research where analysis and synthesis of sounds had been confined to mechanical instrumentation, the 1920s saw the breakthrough of electro-acoustics. Stumpf (1926, 408f.) points already to experiments in radio stations on stereophonic recordings done with several microphones and finds such trends fitting his own concepts, namely, spatialization of sound.

In the 1920s, sound analysis gained new impetus when filtering based on the so-called search tone method (Grützmacher 1927; see also Küpfmüller 1968, 122ff.) allowed spectral decomposition of a time signal. This led first to finding spectra for steady-state sounds. At the same time, transients preceding almost stable sounds such as vowels in speech or tones played on woodwind and bowed string instruments were investigated. The problem in such an approach is that transients lack the clear periodic time structure which governs the steady-state portion of a given sound. Hence, a Fourier analysis that (in principle) assumes strict periodicity of the signal is difficult to conduct if possible at all. Backhaus (1932) who offered a

thorough study of transients (labeled *Ausgleichsvorgänge* then) had the idea to take the period found for the steady-state portion of the signal following the transient portion and see if spectral decomposition could be carried out by tentatively applying the known period length of the stationary signal to the transient part. In many instances, a fair approximation was possible (including manual interpolation of step functions to obtain smoothed curves for the dynamic evolution of partials over time; see Backhaus 1932, 32–34). The results were presented as a family of curves for the partials studied in a 2D amplitude/time diagram. The concept resembles a modern approach, namely the phase vocoder and the 3D-Plots one can obtain for the evolution of partials with appropriate software (such as *sndan*, see Beauchamp 2007). Since the phase vocoder analysis can be understood as a bank of band pass filters tuned to a fundamental frequency (whereby the center frequency of the lowest band pass should closely match f_1 of the sound to be analyzed), a decision as to f_1 or f_0 of the signal has to be made not unlike the considerations Backhaus had outlined. Phase vocoder analysis effects spectral decomposition of a complex harmonic sound into its partials. One option of *sndan* (see Beauchamp 2007) is to plot the amplitude for each partial against time to facilitate the study of onsets and the evolution of spectral energy with time for various partials. For example, looking into the sound produced on a bassoon for the note B_2 ($f_1 \approx 120$ Hz), the 3D plot shows different trajectories for the first six partials within the first 300 ms of sound. The slow rise of most of the partials indicates a rather soft attack for this sound (Fig. 13).

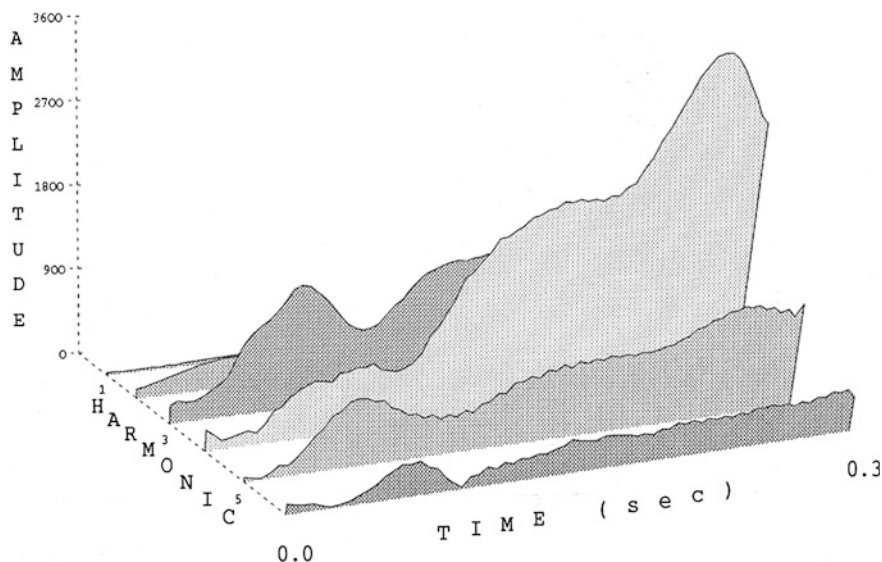


Fig. 13 Phase vocoder analysis, harmonics 1–6, bassoon note B_2 , linear amplitude

Of course, analysis techniques such as the phase vocoder are based on digital signal processing and were not at hand in the 1930s. Acousticians found, however, a clever technique for the study of transients by applying octave-band filters where, because of the wide bandwidth, transient response time of the filter was short enough not to affect the filter output in a significant way.²¹ The filter output per octave and the original signal were recorded on sound film and then plotted as oscillograms whereby different onset times for partials in consecutive octave bands became apparent. Such an analysis was carried out, among others (see Graf 1972; Reuter 1995), on organ flue and reed pipes (Trendelenburg et al. 1936). Studying pipe ranks from the famous organ Arp Schnitger had built for the chapel in the palace of Berlin, in 1706 (cf. Edskes and Vogel 2009, 138–143; the organ was destroyed in WW II), it was found that the onset of partials differs significantly from reed pipes to flue pipes (what can be explained by taking into account characteristics of different generators and regimes of vibration) and also with respect to various pipe geometries and pipe sizes per octave.

Basically the same approach as *Suchtonanalyse* is known also as heterodyne filter analysis (see Fant 1952; Roads and Strawn 1996, 548ff.), which has been adopted in the construction of the analogue sonagraph (designed by Bell Labs) that came into research institutions as a stand-alone hardware unit (built by Kay Elemetrics Co.) in the 1950s and was primarily used for visualization of speech patterns (“visible speech”; cf. Potter et al. 1947; Neppert and Pétursson 1986). The sonagraph offered analysis of a segment of sound either directly recorded into the machine (by microphone) or fed from tape or record player into a line input (mono). The Kay sonagraph became a standard tool in phonetics and, beginning in the 1960s, also for systematic and comparative musicology (see Graf 1972, 1976, 1980; Födermayr 1971; Rösing 1972). The following sonogram displays the same melisma of a female singer from Lebanon that has been under study above (Figs. 2, 3); the sonagram was produced with the (advanced) Kay model 7030 A that offered two exchangeable filter units with four different band filter settings (wide: 300 or 150 Hz; narrow: 45 or 10 Hz).²²

The sonagram clearly shows the modulation and a concentration of spectral energy in bands corresponding to vocal formants (these bands cover ca. 600–800 Hz, 1200–1400 Hz, and from 3.4 to above 4 kHz). The results of the analogue sonographic analysis are well in line with the spectral and formant analysis obtained by digital signal processing (see Figs 2, 3 and 5).

²¹ For the calculation of analogue low pass and band pass filter parameters, see Fant (1952), Küpfmüller (1968).

²² The sonagram was produced, in spring 1978, in the lab of the Kommission für Schallforschung of the Austrian Academy of Sciences at Vienna where the present author was working for a period on invitation of Walter Graf, then head of the Kommission für Schallforschung and its lab. Fant (1952) developed a heterodyne filter that, different from the constant bandwidth filter employed in the sonagraph, offers continuously variable bandwidth along with the continuously variable center frequency.

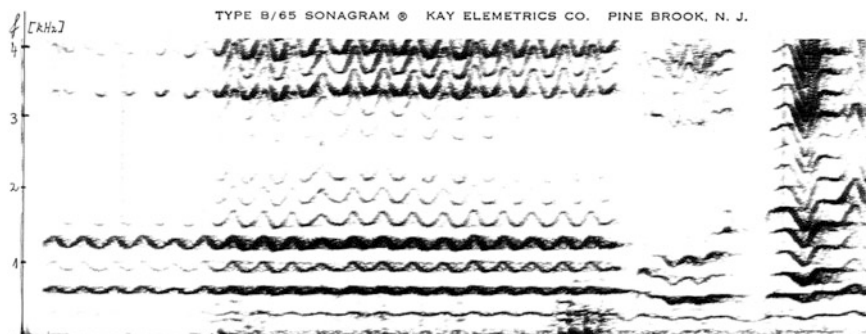


Fig. 14 *Abu Zelf*, melisma, Sonagram 0–4 kHz, 4.8 s of the time signal

Since the analogue sonograph employs a linear frequency scale, one could measure (manually peak-to-peak) the modulation (as in Fig. 14) at the n th partial and then calculate the frequency modulation deviation for the fundamental simply by dividing the shift found at the n th partial by n . Also, one could make use of both wide band and narrow band filters in the same probe and thereby overcome—at least to a certain degree—the limitations known from linear systems (cf. Küpfmüller 1968, Kap. IV), where the relation $df/dt \geq 1$ applies. In regard to band pass filters this means that the filter response time is $\tau = 2\pi/\Delta\omega = 1/\Delta f$, that is, the response takes the longer the narrower the pass band of the filter is chosen. Having the same probe processed by filters of different bandwidth, one could have the advantage of better temporal resolution with the 300/150 Hz filter, and improved frequency resolution by using the 45/10 Hz narrow band (cf. Schneider 1986).

In the 1970s, digital signal analysis based on the Discrete Fourier Transforms (DFT) making use of highly efficient algorithms such as FFT (see Randall 1987; DeFatta et al. 1988) was developing fast. Spectrum analyzers offering DFT/FFT became available though sampling rates, length of the transform for Fourier analysis and, hence, temporal and frequency resolution were still modest (cf. Randall 1987) mostly due to limits in memory needed for storage and processing of signals. In addition to narrow-band spectrum analyzers (such as the B and K 2031 and 2033 models that, by about 1980/81, were found in many labs), also a digital sonograph (Sona-Graph DSP 5500, Kay Elemetrics) was constructed which, like the analogue model, allowed to measure the temporal evolution of spectra and to plot time/frequency representations as spectrograms but also had an option for analysis of two independent signal channels. This machine has been used in the study of transients in organ flue pipes as well as in recorders (Castellengo 1999).

By the end of the 1970s, the Synclavier (and by about 1980 the Synclavier II) became available that, besides its capabilities as a digital synthesizer and high resolution sampler (with sampling up to 100 kHz at 16 bit), offered also state-of-the-art FFT-based spectral analysis including 3D spectral plots and harmonic grid display to check the harmonicity of spectral components (cf. Beurmann and Schneider 1995; Schneider 1997b). Moreover, the Synclavier II offered automated

transcription, whereby sound was directly transformed into western staff notation. We have tried this option along with the spectral analysis and component frequency and level (dB) calculation on, for example, two-part pan pipe music of the 'Are'are (Salomon Islands, recordings made by Hugo Zemp; cf. Schneider 1997b, 393ff.).

By about 1990, powerful workstations (like SGI, SUN, NeXT) had become available, some with A/D and D/A conversion as well as DSP hardware onboard. Also, software for sound analysis and synthesis such as *sndan* (cf. Beauchamp 2007), *Spectro* (Gary Scavone and P. Cook, CCRMA/Stanford), *Sonogram* (Hiroshi Momose, UC Davis), *SuperVP* (IRCAM) as well as versatile platforms for programming (like CSound, CMusic, ESPS +) along with multi-functional math packages (Mathematica, MatLab) opened a completely new age especially for systematic musicologists as well as for composers and musicians interested in acoustics and psychoacoustics, sound synthesis, computer music, etc. [for an overview of tools and fields of application, see comprehensive books edited by Roads and Strawn 1996; Roads et al. 1997; Zölzer 2003; Beauchamp 2007 and the monograph Smith (2007) has supplied]. In regard to sound analysis, careful application of short-time Fourier transforms (STFT) allowed to study transients as well as inharmonic signals. By choosing a very low hop ratio (down to a few samples or even one sample) and, thus, a high percentage of overlap of frames that are processed (and in addition zero padding as needed, depending on the signal structure and transform length; see DeFatta et al. 1988), a quasi-continuous spectral documentation of transients in onsets as well as of modulation processes as observed in various musical instruments became accessible (for examples, see Schneider 1997b, 1998, 2000; Beurmann et al. 1998). Also, spectral envelope estimates by means of LPC or AR algorithms (cf. Marple 1987; Rodet and Schwarz 2007, Schneider and Mores, this volume) allowed close inspection of transient parts of sound. In addition to STFT, wavelet analysis has been employed as a high-resolution time/frequency representation of signals (cf. Mertins 1996, 1999). Since the transient portion of a signal contains most of the information, calculation of fractal dimensions and other methods known from dynamic system analysis (e.g., phase space and limit cycle analysis, Hopf bifurcation) help to identify processes characteristic of the transient part as distinct from the steady-state sound (cf. Bader 2002). One of the motives central to conduct such studies is to understand complex acoustic systems such as musical instruments from their sound as it is radiated from vibrating parts and surfaces. The sound patterns thus are likely to indicate patterns of vibration. Therefore, results obtained from the analysis of sound of real instruments must converge to a high degree with data from vibration analysis and modeling of instruments such as is done with, for example, the finite element and finite difference methodologies (FEM, FDM; see Bader 2005 for his model of the classical guitar and Lau et al. 2010 for a model of a swinging bell as compared to a real bell). In this respect, advanced sound analysis has gained an important role in musical acoustics as well as in areas of psychoacoustics where properties of sound in regard to perception are of interest. To be sure, most of the developments addressed in the present survey (covering, in the main, the era of Helmholtz to

the time when computers and digital signal processing had become widespread) happened within a span of but a hundred years.

4 Summing Up: Continuity and Change

Modern times up to the present seem to be governed by unprecedented speed and ever increasing rates of acceleration. Travel within the 20th century has become faster and faster (from simple automobiles to jet planes), and technological developments have followed one another in ever shorter cycles. A field where the rate of change perhaps is most obvious is information and communication systems including the hardware involved. The computational power of a mainframe of the 1960s or even the 1970s appears small when compared to an average PC in use now. Musical acoustics and sound research have clearly benefited from rapid developments in electronics and in computer technology. Because of the progress in technology and in actual research, the number of publications relating to sound analysis and synthesis is vast. In such a situation, there is a risk that significant portions of previous knowledge will fall into oblivion notwithstanding massive data storage and archiving of relevant publications in digital formats.

Of course, there is always the possibility to look back on earlier achievements in research and technology as has been done for acoustics. In regard to science, historical accounts, rather than being conceived as a plain “narrative” of past endeavours and achievements, must try to reveal topics central to research in a certain field along with shedding light on issues in approach and methodology (including basic mathematical and physical background; see, e.g., Cannon and Dostrovsky 1982; Beyer 1999).

While history conceived as “narrative” can hardly be written without some concept of continuity, Canguilhem (1966/1994) has stressed the importance of discontinuity in science as a driving force for progress. One could also point to the role of changing ‘paradigms’ that supplant or replace one another to underpin the dynamics of science. Though discontinuity is a fact that can be observed in many instances in regard to theories, methodology, and also practical matters, there is also persistence of previous knowledge as well as of problems that are unsolved yet and hence need to be investigated. In order to find improved experimental designs and possibly more appropriate solutions, one has to know about previous research and its outcome (including valid results as well as failures).

The present article and some other publications I have contributed making use of sound analysis and synthesis in one way or another (e.g., Schneider 1997a, b, 2001, 2011) were conceived with the intent to connect the present to the past, and to outline certain developments in regard to both change and continuity in research. While change is often induced by new tools (technical as well as conceptual and methodological) that become available for research, leading to fresh perspectives, in quite many areas a certain continuity can also be observed either because problems resist to be solved completely, or because previous research is

acknowledged as still worth noticing and perhaps calls for a continuation of efforts.

References

- Albersheim, G. (1939). *Zur Psychologie der Ton- und Klangeigenschaften unter Berücksichtigung der ‚Zweikomponententheorie‘ und der Vokalsystematik*. Straßburg: Heitz.
- Arfib, D., Keiler, F., & Zölzer, U. (2002). Source-Filter Processing. In U. Zölzer (Ed.), *DAFX. Digital Audio Effects* (pp. 299–372). Chichester: Wiley.
- Backhaus, H. (1932). Über die Bedeutung der Ausgleichsvorgänge in der Musik. *Zeitschrift für technische Physik*, 13, 31–46.
- Backus, J. (1963). Acoustical investigations of the clarinet. *Sound*, 2, 22–25.
- Bader, R. (2002). *Fraktale Dimensionen, Informationsstrukturen und Mikrorhythmik der Einschwingvorgänge von Musikinstrumenten*. Ph.D. dissertation, University of Hamburg.
- Bader, R. (2005). *Computational Mechanics of the Classical Guitar*. Berlin etc.: Springer.
- Bader, R. (2011). Buddhism, animism, and entertainment in Cambodian melismatic chanting *smot*—history and tonal system. In A. Schneider & A. von Ruschkowski (Eds.), *Systematic Musicology: Empirical and theoretical studies* (pp. 283–305). P. Lang: Frankfurt/M.
- Bathey, B. (2004). Bézier spline modeling of pitch-continuous melodic expression and ornamentation. *Computer Music Journal*, 28, 25–39.
- Beauchamp, J. (2007). Analysis and Synthesis of musical instrument sounds. In J. Beauchamp (Ed.), *Analysis, Synthesis, and Perception of musical sounds* (pp. 1–89). New York: Springer.
- Beurmann, A., & Schneider, A. (1995). Zur akustischen Untersuchung von Volksmusikinstrumenten. *Studia instrumentorum musicae popularis* (Vol. X, pp. 113–121). Stockholm: Musikmuseet.
- Beurmann, A., & Schneider, A. (2008). Acoustics of the harpsichord: a case study. In A. Schneider (Ed.), *Systematic and Comparative Musicology: Concepts, methods, findings* (pp. 241–263). P. Lang: Frankfurt/M.
- Beurmann, A., & Schneider, A. (2009). Acoustics and sound of the harpsichord: another study. In R. Bader (Ed.), *Musical Acoustics, Neurocognition and Psychology of Music* (pp. 57–72). Frankfurt/M: P. Lang.
- Beurmann, A., Schneider, A., & Lauer, E. (1998). Klanguntersuchungen an der Arp-Schnitger-Orgel zu St. Jacobi, Hamburg. *Systematische Musikwissenschaft*, 6, 151–187.
- Beyer, R. (1999). *Sounds of our time. Two hundred years of acoustics*. New York: Springer.
- Boersma, P. (1993). Accurate short-term Analysis of the fundamental frequency and the harmonic-to-noise ratio of a sampled sound. *IFA Proceedings*, 17, 97–110.
- Boersma, P., Weenink D. (2011). *Praat. Doing Phonetics by Computer*. Amsterdam: University of Amsterdam, Institute of Phonetic Sciences. (Praat version 5323).
- Bozkurt, B. (2008). An automatic Pitch analysis method for Turkish Maquam music. *Journal of New Music Research*, 37, 1–13.
- Bozkurt, B., Yarman, O., Karaosmanoğlu, K., & Akkoş, C. (2009). Weighing diverse theoretical Models on Turkish *Maquam* music against pitch measurements: a comparison of peaks automatically derived from frequency histograms with proposed scale tones. *Journal of New Music Research*, 38, 45–70.
- Bregman, A. (1990). *Auditory Scene Analysis*. Cambridge: MIT Press.
- Brown, J. (2007). Fundamental frequency tracking and applications to musical signal analysis. In J. Beauchamp (Ed.), *Analysis, Synthesis, and Perception of musical sounds* (pp. 90–121). New York: Springer.

- Brown, J., & Puckette, M. (1993). A high resolution fundamental frequency determination based on phase changes of the Fourier transform. *Journal of Acoustical Society of America*, 94, 662–667.
- Cañadas Quesada, F. J., Ruiz Reyes, N., Vera Candeas, P., Carabias, J. J., & Maldonado, S. (2010). A multiple-F0 estimation approach based on Gaussian spectral modeling for polyphonic music transcription. *Journal of New Music Research*, 39, 93–107.
- Canguilhem, G. (1966/1994). L'Objet de l'histoire des sciences (= lecture, Montréal 1966). In G. Canguilhem (Ed.), *Études d'histoire et de philosophie des sciences* (pp. 9–23). 7e ed. Paris: Vrin 1994.
- Cannon, J., & Dostrovsky, S. (1982). *The Evolution of Dynamics: Vibration theory from 1687 to 1742*. New York: Springer.
- Cariani, P. (2004). A temporal Model for Pitch multiplicity and tonal Consonance. *Proceedings of International Conference on Music Perception and Cognition (IMPC)*, Evanston, 310–314.
- Castellengo, M. (1999). Analysis of initial transients in flute like instruments. *Acustica*, 85, 387–400.
- Chladni, E. F. F. (1805/1830). *Die Akustik*. Leipzig: Breitkopf & Haertel (2nd ed. 1830).
- Chladni, E. F. F. (1817). *Neue Beyträge zur Akustik*. Leipzig: Breitkopf & Haertel.
- Cohen, H. F. (1984). *Quantifying Music. The Science of music at the first stage of the scientific revolution, 1580–1650*. Dordrecht, Boston: D. Reidel.
- Cohen, L. (1995). *Time-Frequency analysis*. Upper Saddle River: Prentice Hall.
- Dahlback, K. (1958). *New Methods in vocal folk music research*. Oslo: University Press.
- Dalmont, J. P., Gilbert, J., & Kergomard, J. (2000). Reed Instruments, from small to large amplitude periodic oscillations and the Helmholtz motion analogy. *Acustica*, 86, 671–684.
- de Boer, E. (1976). On the “Residue” and Auditory Pitch Perception. In W. D. Keidel & W. D. Neff (Eds.), *Handbook of Sensory Physiology* (pp. 479–583). Berlin: Springer.
- de Cheveigné, A., & Kawahara, H. (2002). YIN, a fundamental frequency estimator for speech and music. *Journal of the Acoustical Society of America*, 111, 1917–1930.
- de Cheveigné, A. (2005). Pitch perception models. In Chr. Plack, A. Oxenham, R. Fay, A. Popper (Eds.), *Pitch. Neural Coding and Perception* (pp. 169–233). New York: Springer.
- DeFatta, D., Lucas, J., & Hodgkiss, W. (1988). *Digital Signal Processing. A system design approach*. New York: Wiley.
- Edison, T. (1888). The perfected Phonograph. *The North American Review*, 146(379), 641–650.
- Edskes, C., & Vogel, H. (2009). *Arp Schnitger und sein Werk*. Bremen: Hauschild.
- Ehret, G. (1997). The auditory midbrain, a “shunting yard” of acoustical information processing. In G. Ehret & R. Romand (Eds.), *The Central Auditory System* (pp. 259–316). Oxford: Oxford University Press.
- Elschek, O. (1979). *Melographische Interpretationscharakteristika von Flötenmusik. Studia instr. mus. pop.* VI (pp. 43–58), Stockholm: Musikhist. Museet.
- Elschek, O. (2006). *Fujara. The Slovak Queen of European Flutes*. Bratislava: Hudobné Centrum.
- Fant, G. (1952). *The Heterodyne Filter*. Göteborg/Stockholm: Elander/Lindstahl.
- Fant, G. (1960). *Acoustic theory of speech production. With calculations based on x-ray studies of Russian articulations*. The Hague: Mouton.
- Feld, S. (1990). *Sound and Sentiment: Birds, weeping, poetics, and song in Kaluli expression* (2nd ed.). Philadelphia: University of Pennsylvania Press.
- Feld, S., Kirgegaard, A. (2010). Entangled Complicities in the prehistory of ‘World Music’: Poul Rovsing Olsen and Jean Jenkins encounter Brian Eno and David Byrne in the Bush of Ghosts. *Popular Musicology Online*. www.popular-musicology-online.com/issues/04/feld.html. (retrieved 30th of August, 2012).
- Filip, M. (1969). Envelope periodicity detection. *Journal of the Acoustical Society of America*, 45, 719–732.
- Filip, M. (1970). Frekvenčné merania a tónová sústava [Frequency measurement and tonal system] (pp. 50–85), *Nové Cesty Hudbe*, 2, Praha.

- Filip, M. (1978). Acoustic measurements as auxiliary methods in ethnomusicology. *Musicologica Slovaca*, 7, 77–87.
- Födermayr, F. (1971). *Zur gesanglichen Stimmgebung in der außereuropäischen Musik* (Vol. 2). Wien: Stiglmayr.
- Fricke, J. (1993). Die Wechselwirkung von Mensch und Musik im Zusammenspiel von Physik und Physiologie. In B. Enders & St. Hanheide (Eds.), *Neue Musiktechnologie* (pp. 169–196). London: Schott.
- Graf, W. (1972). Musikalische Klangforschung. *Acta Musicologica*, 44, 31–78.
- Graf, W. (1976). Zum Klang der Stainer-Geigen. In *Festschrift Walter Senn zum 70. Geburtstag* (pp. 98–117), hrsg. von E. Egg u.a., München: Katznbichler.
- Graf, W. (1980). *Vergleichende Musikwissenschaft. Ausgewählte Aufsätze*. Hrsg. von F. Födermayr. Stiglmayr: Wien-Föhrenau.
- Grützmaker, M. (1927). Eine neue Methode der Klanganalyse. *Elektrische Nachrichtentechnik (ENT)*, 4(12), 533–545.
- Grützmaker, M., & Lottemoser, W. (1937). Über ein Verfahren zur trägheitsfreien Aufzeichnung von Melodiekurven. *Akustische Zeitschrift*, 2, 242–248.
- Hartmann, W. (1998). *Signals, Sound, and Sensation*. New York: AIP/Springer.
- Hekland, F. (2001). *Automatic Music transcription using autoregressive frequency estimation*. Project paper. Toulouse: ENSEEIHT.
- Helmholtz, H. (1863/1870/1896). *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik*. Braunschweig: Vieweg (3rd ed. 1870, 5th ed. 1896).
- Helmholtz, H. von (1857/1896). Über die physiologischen Ursachen der musikalischen Harmonien (Vortrag, Bonn 1857). In H. von Helmholtz (Ed.), *Vorträge und Reden* (pp. 119–155). 4. Aufl. Braunschweig: Vieweg 1896, Bd 1.
- Hermann, L. (1889). Phonographische Untersuchungen. [Pflügers]. *Archiv für die gesamte Physiologie*, 45, 582–592.
- Hermann, L. (1891). Bemerkungen zur Vocalfrage. *Archiv für die gesamte Physiologie*, 48, 181–194.
- Hermann, L. (1893). Phonographische Untersuchungen. IV: Untersuchungen mittels des neuen Edison'schen Phonographen. *Archiv für die gesamte Physiologie*, 53, 1–51.
- Hermann, L. (1894). Phonographische Untersuchungen. VI. Nachtrag zur Untersuchung der Vocalcurven. (Nach Versuchen in Gemeinschaft mit Dr. F. Matthias und stud. med. Alfred Ehrhardt.) *Archiv für die gesamte Physiologie*, 58, 264–279.
- Hermann, L. (1911). Neue Beiträge zur Lehre von den Vokalen und ihrer Entstehung. *Archiv für die gesamte Physiologie*, 141, 1–62.
- Hermann, L. (und H. Hirschfeld) (1895). Weitere Untersuchungen über das Wesen der Vocale. *Archiv für die gesamte Physiologie*, 61, 169–209.
- Herrmann, E. (1908). Über die Klangfarbe einiger Orchesterinstrumente und ihre Analyse. Diss. Königsberg 1908. In *Beiträge zur Physiologie und Pathologie. Festschrift zum 70. Geburtstag L. Hermann ...gewidmet* (pp. 59–105), hrsg. von O. Weiss, Stuttgart: Union Deutsche Verlagsges.
- Hesse, H. P. (1972). *Die Wahrnehmung von Tonhöhe und Klangfarbe als Problem der Hörtheorie*. Köln: Volk/Gerig.
- Ingle, V., & Proakis, J. (2000). *Digital Signal Processing using MATLAB*. Pacific Grove ets.: Brooks/Cole.
- Keidel, W. D. (Ed.). (1975). *Physiologie des Gehörs. Akustische Informationsverarbeitung*. Stuttgart: Thieme.
- Kergomard, J., Ollivier, S., & Gilbert, J. (2000). Calculation of the Spectrum of self-sustained oscillators using a variable truncation method: application to cylindrical reed instruments. *Acustica*, 86, 685–703.
- Klapuri, A. (2004). Automatic music transcription as we know it today. *Journal of New Music Research*, 33, 269–282.
- Klapuri, A., & Davy, M. (Eds.). (2006). *Signal Processing Methods for music transcription*. New York: Springer.

- Kreichgauer, A. (1932). *Über Maßbestimmungen freier Intonationen*. Berlin: Phil. Diss.
- Krueger, F. (1907). Beziehungen der experimentellen Phonetik zur Psychologie. *Bericht über den II. Kongress für exp. Psychol. Würzburg 1906* (pp. 58–122). Leipzig: Barth.
- Küpfmüller, K. (1968). *Die Systemtheorie der elektrischen Nachrichtenübertragung*. 3. Aufl. Stuttgart: Hirzel.
- Lau, B., Bader, R., Schneider, A., & Wriggers, P. (2010). Finite-Element transient calculation of a bell struck by its clapper. In R. Bader, Chr. Neuhaus, & U. Morgenstern (Eds.), *Concepts, Experiments, and Fieldwork: Studies in Systematic Musicology and Ethnomusicology* (pp. 137–156). Frankfurt/M.: P. Lang.
- Licklider, J. C. (1951). A duplex theory of pitch perception. *Experientia*, 7, 128–134.
- Licklider, J. C. (1956). Auditory Frequency Analysis. In C. Cherry (Ed.), *Information theory* (pp. 253–268). London: Butterworth.
- Lottermoser, W. (1976/1977). Frequenzschwankungen bei musikalischen Klängen. *Acustica*, 36, 138–146.
- Marple, S. L. (1987). *Digital Spectral Analysis*. Englewood Cliffs, NJ: Prentice Hall.
- McAulay, R., Quatieri, T. (1986). Speech Analysis/Synthesis based on a sinusoidal representation. *IEEE Transactions on Acoustics, Speech, and Signal Processing* Vol. ASSP-34, 34(4), 744–754.
- Meddis, R., & O'Mard, L. (1997). A unitary Model of pitch perception. *Journal of the Acoustical Society of America*, 102, 1811–1820.
- Meddis, R., & O'Mard, L. (2006). Virtual pitch in a computational physiological model. *Journal of the Acoustical Society of America*, 120, 3861–3869.
- Meissner, G. (1907). Klangaufnahmen an Blasinstrumenten, eine Grundlage für das Verständnis der menschlichen Stimme. Nachgelassenes Manuscript ...hrsg. durch Richard Wachsmuth. [Pflügers]. *Archiv für die gesamte Physiologie*, 116, 543–599.
- Mertens, P. H. (1975). *Die Schumannschen Klangfarbengesetze und ihre Bedeutung für die Übertragung von Sprache und Musik*. Frankfurt/M: Bochinsky.
- Mertins, A. (1996). *Signaltheorie*. Stuttgart: Teubner.
- Mertins, A. (1999). *Signal Analysis*. Chichester: Wiley.
- Metfessel, M. (1928). *Phonophotography in Folk Music. American Negro songs in new notation*. Chapel Hill: University of North Carolina Press.
- Meyer, E., & Guicking, D. (1974). *Schwingungslehre*. Braunschweig: Vieweg.
- Mores, R. (2010). Vowel quality in violin sounds. In R. Bader, Chr. Neuhaus, & U. Morgenstern (Eds.), *Concepts, Experiments, and Fieldwork: Studies in Systematic Musicology and Ethnomusicology* (pp. 113–135). Frankfurt/M.: P. Lang.
- Nagel, G. (2007). Sprengstoff- und Fusionsforschung an der Berliner Universität: Erich Schumann und das II. Physikalische Institut. In R. Karlsch & H. Petermann (Eds.), *Für und Wider 'Hitlers Bombe'. Studien zur Atomforschung in Deutschland* (pp. 229–260). Münster: Waxmann.
- Nelken, I. (2002). Feature detection by the auditory cortex. In D. Oertel, R. Fay, & A. Popper (Eds.), *Integrative Functions in the Mammalian auditory Pathway* (pp. 358–416). New York: Springer.
- Neppert, J., & Pétursson, M. (1986). *Elemente einer akustischen Phonetik* (2nd ed.). Hamburg: Buske.
- Obata, J., & Kobayashi, R. (1937). A direct-reading pitch recorder and its application to music and speech. *Journal of the Acoustical Society of America*, 9, 156–161.
- Opelt, F. W. (1852). *Allgemeine Theorie der Musik auf den Rhythmus der Klangwellenpulse gegründet*. Leipzig: Barth.
- Owen, T. (1974). Applying the Melograph to "Parker's Mood". *Selected Reports in Ethnomusicology* (pp. 167–175), Vol. II, 1. Los Angeles: UCLA.
- Paiva, R. P., Mendes, T., & Cardoso, A. (2008). From Pitches to notes: creation and segmentation of pitch tracks for melody detection in polyphonic audio. *Journal of New Music Research*, 37, 185–205.

- Panconcelli-Calzia, G. (1941/1994). *Geschichtszahlen der Phonetik. 3000 Jahre Phonetik*. Hamburg: Hansischer Gildenverlag 1941; Repr. Amsterdam: Benjamins 1994.
- Pipping, H. (1890). Zur Klangfarbe der gesungenen Vocale. *Zeitschrift für Biologie* 27, 1ff., 433ff.
- Pipping, H. (1894). *Über die Theorie der Vokale*. Helsingfors: Soc. Litterariae Fennicae (=Acta Soc. Scient. Fenn. XX, no. 11).
- Potter, R., Kopp, G., & Green, H. (1947). *Visible Speech*. New York: van Nostrand.
- Randall, R. B. (1987). *Frequency Analysis* (3rd ed.). Naerum: Bruel & Kjaer.
- Reinecke, H. P. (2003). Hermann von Helmholtz, Carl Stumpf und die Folgen. Von der Akustik zur Tonpsychologie. Über ein Kapitel Wissenschaftsgeschichte in Berlin. In M. Kaiser-El-Safti & M. Ballod (Eds.), *Musik und Sprache. Zur Phänomenologie von Carl Stumpf* (pp. 185–197). Würzburg: Königshausen & Neumann.
- Reuter, C. (1995). *Der Einschwingvorgang nichtperkussiver Musikinstrumente*. Frankfurt/M: P. Lang.
- Reuter, C. (1996). *Die auditive Diskrimination von Orchesterinstrumenten. Verschmelzung und Heraushörbarkeit von Instrumentalklangfarben im Ensemblespiel*. Frankfurt/M: P. Lang.
- Roads, C., Strawn, J., et al. (1996). *The Computer Music Tutorial*. Cambridge: MIT Press.
- Roads, C., Pope, S. T., Piccialli, A., & de Poli, G. (Eds.). (1997). *Musical Signal Processing*. Lisse etc.: Swets & Zeitlinger.
- Rodet, X., & Schwarz, D. (2007). Spectral Envelopes and additive + residual analysis/synthesis. In J. Beauchamp (Ed.), *Analysis, Synthesis, and Perception of musical sounds* (pp. 174–227). New York: Springer.
- Rösing, H. (1972). *Die Bedeutung der Klangfarbe in traditioneller und elektronischer Musik. Eine sonographische Untersuchung*. München: Katzbichler.
- Sauveur, J. (1701). Système général des intervalles des sons, et son application à tous les systèmes et à tous les instruments de musique. *Histoire de L'Académie Royale des Sciences. Année 1701, Mémoires de Mathématique & de Physique*, 297–364.
- Schneider, A. (1986). Tonsystem und Intonation. *Hamburger Jahrbuch für Musikwissenschaft*, 9, 153–199.
- Schneider, A. (1987). Musik, Sound, Sprache, Schrift: Transkription und Notation in der Vergleichenden Musikwissenschaft und Musikethnologie. *Zeitschrift für Semiotik*, 9, 317–343.
- Schneider, A. (1997a). “Verschmelzung” Tonal Fusion, and Consonance: Carl Stumpf revisited. In M. Leman (Ed.), *Musik, Gestalt, and Computing. Studies in Cognitive and Systematic Musicology* (pp. 117–143). Berlin: Springer.
- Schneider, A. (1997b). *Tonhöhe–Skala–Klang. Akustische, tonometrische und psychoakustische Studien auf vergleichender Grundlage*. Bonn: Orpheus-Verlag für syst. Musikwiss.
- Schneider, A. (1998). Klanganalysen bei Aerophonen der Volksmusik. In F. Födermary & L. Burlas (Eds.), *Ethnologische, Historische und Systematische Musikwissenschaft. Oskár Elschek zum 65. Geburtstag* (pp. 51–80). Bratislava: ASCO
- Schneider, A. (2000). Inharmonic Sounds: Implications as to «Pitch», «Timbre» and «Consonance». *Journal of New Music Research*, 29, 275–301.
- Schneider, A. (2001). Sound, Pitch, and Scale: from ‘tone measurements’ to sonological analysis in ethnomusicology. *Ethnomusicology*, 45, 489–519.
- Schneider, A. (2011). Music Theory: Speculation, Reasoning, Experience. A Perspective from Systematic Musicology. In T. Janz & J. Ph. Sprick (Eds.), *Musiktheorie | Musikwissenschaft. Geschichte—Methoden—Perspektiven* (pp. 53–97). New York: Olms.
- Schneider, A., & Frieler, L. (2009). Perception of harmonic and inharmonic sounds: results from ear models. In S. Ystad, R. Kronland-Martinet, & K. Jensen (Eds.), *Computer Music Modeling and Retrieval. Genesis of meaning in sound and music* (pp. 18–44). Berlin: Springer.
- Schneider, A., von Busch, R., & Schmidt, L. (2001). Klanganalysen an Arp Schnitger-Orgeln. In N. Ristow, W. Sandberger, & D. Schröder (Eds.), *«Critica Musica.» Studien zum 17. und 18. Jahrhundert* (pp. 247–270). Stuttgart: Metzler.
- Schumann, K. E. (1925). *Akustik*. Breslau: Hirt.

- Scripture, E. W. (1906). *Research in Experimental Phonetics. The Study of Speech Curves*. Washington, D.C.: Carnegie Institution.
- Scripture, E. W. (1927). *Anwendung der graphischen Methode auf Sprache und Gesang*. Leipzig: J. Barth.
- Seeger, C. (1951). An instantaneous music notator. *Journal of the International Folk Music Council*, 3, 103–106.
- Slawson, W. (1985). *Sound Color*. Berkeley: University of California Press.
- Smith, J. O. (2007). Introduction to digital filters with audio applications. Stanford: CCRMA (online book available at <https://ccrma.stanford.edu/~jos/filters/>).
- Stern, W. (1902). Der Tonvariator. *Zeitschrift für Psychologie*, 30, 422–432.
- Stevens, S. S. (1934). The Volume and Intensity of Sounds. *The American Journal of Psychology*, 46, 397–408.
- Stumpf, C. (1890). *Tonpsychologie*. Bd II. Leipzig: J. Barth.
- Stumpf, C. (1898). *Konsonanz und Dissonanz*. Leipzig: J. Barth.
- Stumpf, C. (1926). *Die Sprachlaute Experimentell-phonetische Untersuchungen nebst einem Anhang über Instrumentalklänge*. Berlin: Springer.
- Sundberg, J. (1997). *Die Wissenschaft von der Singstimme*. Bonn: Orpheus-Verlag für syst. Musikwiss.
- Terhardt, E. (1998). *Akustische Kommunikation*. Berlin: Springer.
- Tjernlund, P., Sundberg, J., Fransson, F. (1972). Grundfrequenzmessungen an schwedischen Kernspaltflöten. *Studia instr. mus. pop.* II, Stockholm, (pp. 77–96).
- Tramo, M., Cariani, P., Delgutte, B., & Braidia, L. (2001). Neurobiological Foundations for the theory of harmony in western tonal music. *Annals of the New York Academy of Sciences*, 930, 92–116.
- Trendelenburg, F., Thienhaus, E., & Franz, E. (1936). Klangeinsätze an der Orgel. *Akustische Zeitschrift*, 1, 59–76.
- Vierling, O. (1936). Der Formantbegriff. *Annalen der Physik*, 5, Folge, Bd 26, 219–232.
- Voigt, W. (1975). *Untersuchungen zur Formantbildung in Klängen von Fagott und Dulzianen*. Regensburg: Bosse.
- von Quanten, E. (1875). Einige Bemerkungen zur Helmholtz'schen Vocalehre. [Poggendorfs]. *Annalen der Physik*, 230, 522–552.
- Weber, G. (1817/1824/1830). *Versuch einer geordneten Theorie der Tonsetzkunst*. Bd 1, Mainz: Schott (2nd ed. 1824, 3rd ed. 1830).
- Wiener, N. (1961). *Cybernetics or control and communication in the animal and in the machine* (2nd ed.). New York: MIT Press.
- Willis, R. (1830). On the vowel sounds, and on the reed organ-pipes. *Transactions of the Cambridge Philosophical Society*, 3, 229–268.
- Winckel, F. (1960). *Phänomene des musikalischen Hörens*. Berlin: Hesse.
- Zölzer, U. (Ed.). (2003). *DAFX. Digital Audio Effects*. Chichester: Wiley.