32. Perception of *Timbre* and *Sound Color*

Albrecht Schneider

This chapter deals with perception of *timbre* or *sound color*. Both concepts can be distinguished in regard to terminology as well as to their historical and factual background even though both relate to some common features for which an objective (acoustic) basis exists. Sections of this chapter review in brief developments in (traditional and electronic) musical instruments as well as in research on timbre and sound color. A subchapter on sensation and perception of timbre offers a retrospective on classical concepts of tone color or sound color and reviews some modern approaches from Schaeffer's *objet sonore* to semantic differentials and multidimensional scaling. Taking a functional approach, acoustical features (such as transients and modulation) and perceptual attributes of timbre as well as interrelations between *pitch* and *timbre* are discussed. In a final section, fundamentals of sound segregation and auditory streaming are outlined. For most of the phenomena covered in this chapter, examples are provided including sound analyses obtained with signal processing methods.

32.1 Timbre and Sound Color: Basic Features

32.1.1 Terminology: *Timbre* and *Sound Color*

Timbre is a French word that denotes a stamp (e.g., *timbre fiscal* = revenue stamp), a brand, a sound or a sound color. The French Encyclopédie has an entry *timbre* (T. 16, 1765, 333) that contains several musically relevant annotations, namely *timbre* as referring to snare strings on a drum skin, timbre as the resonant state of a bell, timbre of a voice or of a musical instrument. In his article on *son* (tone, sound; Encyclopédie, T. 15, 1765, 345; for historical aspects see [32.1, 2]), Rousseau had used the term timbre as covering a third property of sounds besides tone height (*le degrè d'élevation entre le grave & l'aigu*) and intensity expressed as the degree *de véhémence entre le fort & le foible*. Timbre then is the quality of a sound that results from an evaluation in regard to dullness – shrillness or softness – brightness (*du sourd à l'éclatant, ou de l'aigu de doux*; the scaling of timbre thus is from sourd and doux to éclantant and aigu). The term timbre as it occurs in textbooks on orchestration [32.3, 4] denotes an integral quality of sound as produced by, and attributed to, certain instruments. These were classified, first by *Mahillon* [32.5–8] and then by *Hornbostel* and *Sachs* [32.9], according to physical principles of sound production, in the first place. The classification offered by Mahillon, based on the huge collection of instruments housed in the Brussels Conservatory of music, comprised:

1. Autophones
2. Instruments à membranes
3. Instruments à vent
4. Instruments à cordes.

R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, https://doi.org/10.1007/978-3-662-55004-5_32

Hornbostel and Sachs have the same four classes (labeled idiophone, membranophone, chordophone, aerophone), however, in a different order (and with significant differences of grouping within each class) that reflects the physical principles involved more clearly. In general, idiophones viewed as vibrating bodies consist of three-dimensional structures (rods, bars, plates, shells), membranes such as drum skins are – ideally – two-dimensional (neglecting their thickness), and thin strings have been treated as one-dimensional in the seminal work of Bernoulli and Euler so that longitudinal, transversal and torsional vibration can be covered by second-order differential equations [32.10, 11]. Finally, aerophones such as flutes, for the lack of shearing forces between molecules in air columns, only allow longitudinal vibration.

Referring to the principles and materials of sound production, *Gevaert* [32.3, p. 5] for example states that the timbre of wind instruments (instruments à vent) is determined by the geometry (length, diameter, bore profile) of the tube, on the one hand, and by the mechanism by which the air inside the tube is set to vibration, on the other. Gevaert correctly points to a regular sequence of pulses (*battements*) necessary for making the air column vibrate, and he also mentions three basic types of pulse generator (edge tone, valves formed by either beating reeds or the lips pressed into a mouthpiece) as are used for sound production in flutes, reeds, and brass instruments respectively. Further, he attributes the *timbre moelleux* (soft timbre) of the French horn to the conical shape of the mouthpiece. *Gevaert* [32.3, p. 18] found that bowed strings are *the soul of instrumental music* since they have a *timbre pénétrant et riche*. *Berlioz* [32.4, p. 21] judged that the *sons harmoniques* of the lowest (fourth) string of a violin *ont quelque chose du timbre du Flûte; ils sont préférable pour chanter une mélodie lente*. The point of interest here is that timbre is regarded as a unique and integral quality (e.g., *timbre du Flûte*), though this may be limited to a certain register or to tones played on a certain string. In this respect, *Forsyth* [32.12, p. 480], in his textbook on orchestration, argues the D (second) string of the cello, *of all the soft, silky sounds in the orchestra, it is the softest and silkiest*.

The unique and consistent timbre attributed to certain instruments or to their parts such as individual strings or pipe ranks (as in organs) is expressed, in the English language, by terms like tone quality, tone color or sound color. Likewise, these terms are used in German (*Klangfarbe*, *Tonfarbe*). Timbre is often used synonymously for *sound color* though there seem to be some differences between the phenomena covered by these terms in that *sound color* predominantly refers to a certain spectral structure while timbre, at least in more recent research, can cover spectral as well as temporal aspects (Sect. 32.1.2). Fundamental to both terms is the experience that sounds can have a distinct sensory quality that, though perhaps not independent of other qualities, in particular pitch and loudness, cannot be accounted for as a *function* of either pitch or loudness alone or as a simple combination of both. Consequently, there seems to be information encoded in the sound structure that gives rise to sensations of *timbre* in addition to information pertaining to *pitch* and to *loudness*.

32.1.2 Objective Basis of *Timbre* and *Sound Color*

In line with the realist and causal perspective taken in Chaps. 30 and 31, timbre and sound color are regarded as sensory experiences deeply rooted in natural foundations. These are physical, on the one hand, and anatomical as well as physiological, on the other. In an evolutionary perspective, various animal species (from insects and amphibians to birds and mammals) show remarkable diversity in regard to organs suited to perform sound production as well as sound perception (articles in [32.13–15]). Many vocalizations serve to communicate information (*Tembrock* [32.16] and articles in *Witzany* [32.17]). The degree of structural and functional complexity reached in birdsong [32.18, 19] and in whale songs [32.20] is particularly striking in regard to the large and diverse song repertoire of certain species involving learning and memory as cognitive capacities as well as communication networks operated by two or more members of a certain species. In addition, interspecies sound communication is known from bioacoustical observations. The songs of birds and whales comprise complex sound patterns (take, for example, songs of the nightingale and of the humpback whale), which vary considerably in spectral content over time. In this respect, it can be said that these sounds make use of timbral qualities as a means of communication. There are parallels in human speech and singing in that different phonation (resulting in different spectral energy distribution and formant structure) can convey different emotional states as well as the intentions of a person speaking or singing [32.21].

Sound production and sound perception in animals (including humans) is based on acoustical principles [32.21–23] even though these are implemented into organic biosystems. In fact, one can give a description of the vocal tract of humans in terms of anatomical structure and functional aspects of muscles, nerves, etc. in the phonation process. Further, one may go into the acoustics of phonation in terms of generator and resonator geometry, air flow and pulse sequence

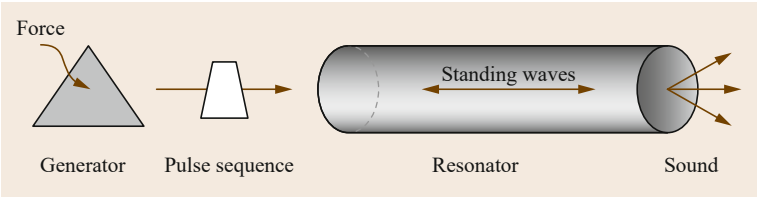


Fig. 32.1 Model of an aerophone (pulse generator coupled to a tube-like resonator)

generation at the vocal folds as well as resonance phenomena taking place in the mouth cavity [32.21, 24] etc. Quite obviously, there are also parallels between the structure and function of the human singing voice and a host of musical instruments classed as aerophones. *Gevaert* [32.3] rightly attributed the timbre of wind instruments to a combination of a valve-like pulse generator and a resonator in which an air column is set to vibration. Taking the general model of such a generator coupled to a resonator, the scheme in Fig. 32.1 can be drawn.

In this basic model, a source (a person breathing air) drives a generator by supplying a force, which in this case is air flowing at a certain speed and with a certain pressure. The flow is periodically interrupted either by a small jet of air being bend inward and outward of a sharp edge (e.g., the labium of an organ flue pipe), or by valves that are (partially or completely) opened and shut to produce pulses of air released into the resonator. A single reed (as in the clarinet) or double reeds (as in the oboe) or the lips pressed into a mouthpiece (as in the trumpet, trombone, French horn, etc.) can serve as the valve in the generator. The sequence of pulses traveling through the length of the resonator (a cylindrical or conical tube) partially is reflected at the end opposite the generator so that standing waves are formed inside the tube where natural modes of vibration are excited. Without going into details, which are intricate because the generator parameters include quite many nonlinearities [32.11, 25], one can see that the behavior of such a generator-plus-resonator system depends on the geometry of its parts, on the input parameters (input impedance of the valve, blowing pressure and speed, air flow through the valve, input impedance of the tube, etc.) and on the coupling of the generator to the resonator. The interaction between the two maintains the regeneration cycle needed for continuant tones. In regard to the geometry of resonators, it has been demonstrated [32.26] that a resonator treated as a Bessel horn produces a series of natural mode frequencies so that higher mode frequencies are multiples of the lowest only if the Bessel function J^x yields either $x = 0$ (cylinder) or $x = 2$ (conical tube). The clarinet has a cylindrical tube almost closed at one end by the valve so that the resonator predominantly responds to odd harmonics; the clarinet overblows into the twelfth, which is the fifth above the octave (the third harmonic [32.27, p. 115–125]), while the oboe (where the bore is close to a cone slightly truncated near the reed generator) overblows into the octave. For the tones in between, finger holes are provided on both instruments. Since also the number of modes excited in the resonator differs between woodwinds (for the same pitch played with similar force of excitation applied), different spectral energy distributions result that are perceived as differences in *sound color* or timbre. For instance, sounds recorded from a number of woodwinds playing the same note (C_4) with moderate force of excitation have spectra that differ in the number and strength of partials (Fig. 32.2).

The sounds analyzed with a phase vocoder algorithm (equivalent to putting the sounds through a bank of band pass filters tuned to f_1 of the sounds [32.28]) are samples of the following instruments (left to right): bassoon 1, bassoon 2, bass clarinet, clarinet, oboe, concert flute. The spectral centroid (Sect. 30.2) for these sounds varies from ≈ 680 Hz (bassoon 1) and 880 Hz (bassoon 2) to 2.4 kHz (bass clarinet) and 2.55 kHz (clarinet); the centroid for the oboe sound is 1.42 kHz while, for the flute, the centroid is identical with f_1 (261.6 Hz). Thus, for a performer or listener the sounds in question differ significantly in brightness when playing the same note. What distinguishes tones produced by different instruments such as the bassoon, the clarinet, the oboe, etc. in the steady-state portion of sound after the transient part is the shape of the spectral envelope and, thereby, the distribution of spectral energy covered by the envelope. The spectra of wind instruments controlled by a valve as pulse generator have been said to approach a cyclic structure with maxima that can be interpreted as formants [32.29–31]. A condition necessary for a perfect cyclic spectrum would be a series of rectangular pulses with a duty cycle of τ/T (τ = pulse width and T = pulse period) that yields small integer ratios; amplitudes of partials then conform to a sinc function $(\sin x)/x$ where zeros are at $n\tau/T = 1, 2, 3, \ldots, k$ [32.32, p. 40 f.] (Fig. 32.3).

Spectra approaching a cyclic structure to some degree can be recorded from reed instruments such as the oboe; in Fig. 32.4, the spectrum of the tone C_4 played (*mf*) on a baroque oboe is shown. In addition, the spectral envelope calculated from a formant filter analysis

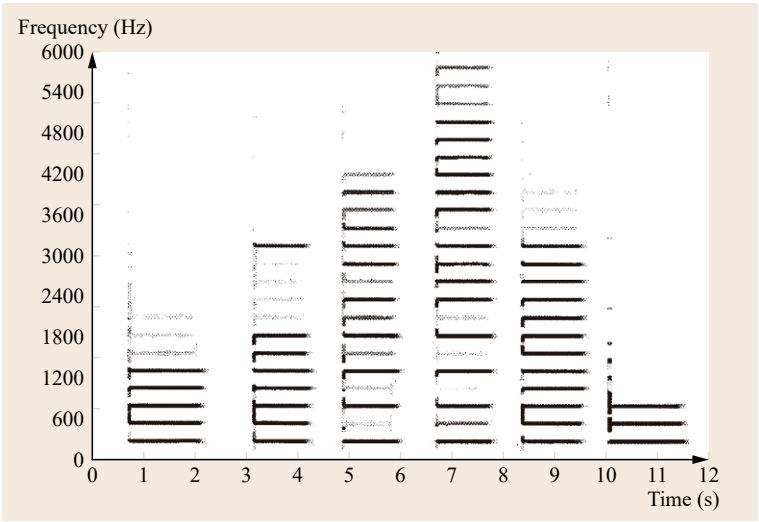


Fig. 32.2 Spectra of various woodwinds (from left to right: bassoon 1, bassoon 2, bass clarinet, clarinet, oboe, flute) all playing the note C_4 (phase vocoder analysis, base frequency = 261.6 Hz). Relative amplitude of partials indicated by grayscale (*white* = low, *black* = high amplitude)

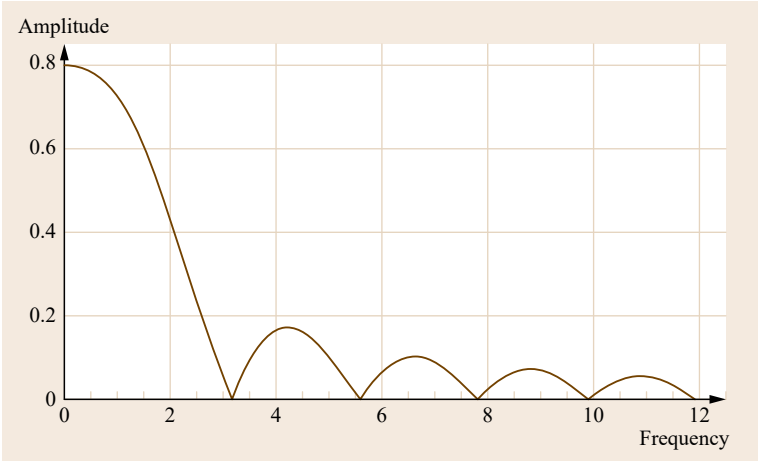


Fig. 32.3 Sinc function as a model for the envelope of a cyclic spectrum

is plotted in the same graph. One can see three groups of partials at $\approx 0-1.8$ kHz, 1.8–3.6 kHz, 3.6–6.6 kHz separated by relative minima; these groups are covered by peaks of the formant filter envelope. Also, a formant analysis performed with the Burg algorithm [32.33] yields several tracks of formant frequencies as a function of time. In this respect, one may assign a formant-like quality to this sound (as well as to sounds from other reed instruments [32.34]). Even closer approximations to a cyclic spectrum can be observed in organ reed pipes [32.35] and also in sounds recorded from plucked strings in a harpsichord where in particular the string velocity reflects the pulse train and the spectrum of the string velocity consequently is fairly periodic [32.36]. Hence, the pulse generator imposes the shape of a spectrum on the resonator, which is fixed in geometry, in organ reed pipes and in the harpsichord, while in reed instruments such as the oboe the length of the air column vibrating in the bore can be modified by means of finger holes.

The generator-plus-resonator model can also be viewed as a generator producing a source signal fed into a filter, that is, into the resonator (for source-filter processing see [32.37, 38]). In terms of linear systems [32.39, Chap. 5] and [32.40, Chap. 9], a filter of low-pass or bandpass characteristic has a certain frequency response as well as an impulse response; the transfer function $H(\omega)$ of a filter determines the amplitudes and phases of the frequency components in the output spectrum relative to the input spectrum. Relating bandwidth to response time, the filter response time τ of a symmetric bandpass to an input signal with rapid onset is

$$\tau = \frac{2\pi}{\Delta\omega} = \frac{1}{\Delta f} .$$

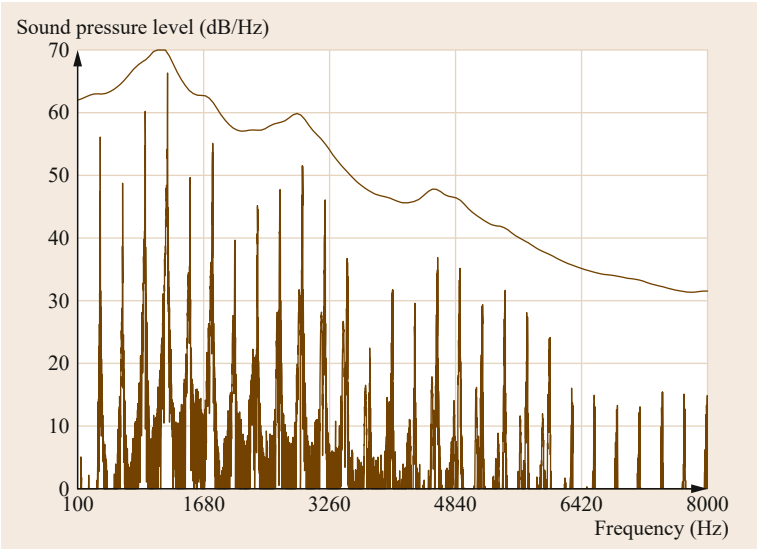


Fig. 32.4 Baroque oboe, spectrum of tone C_4 (*mf*) and formant filter envelope

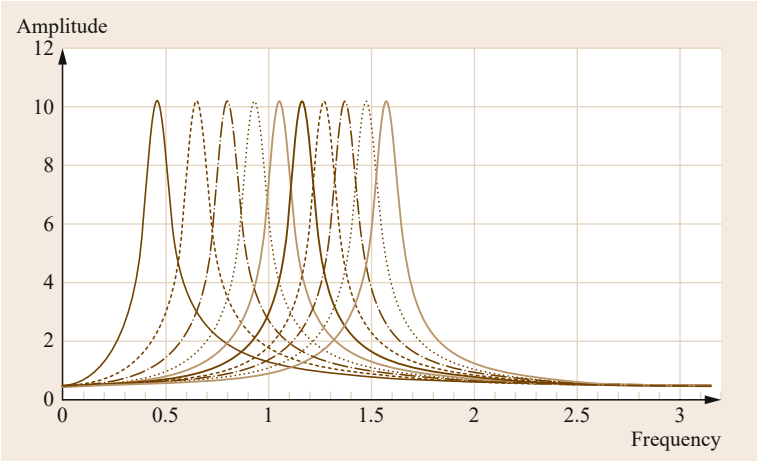


Fig. 32.5 Bank of bandpass filters as analogue of resonance peaks in instruments

Hence the response time is the inverse of the bandwidth. The transfer function of the resonator can be modeled by taking the resonances of the tube as the center frequencies of a chain of bandpass filters (Fig. 32.5).

If we conceive of a digital filter where $x(k)$ is the input signal and $y(k)$ is the output signal, convolving the input signal with the impulse response of the filter, $h(k)$, yields the output sequence $y(k) = x(k) \times h(k)$. The poles of the filter in the complex z-plane are equivalent to the resonance maxima in the resonator (e.g., the tube of an organ flue pipe). It is in physical modeling of musical instruments that such concepts have been implemented in signal processing codes.

In regard to the temporal behavior of the system, it takes some time before a pulse sequence fed into a tube resonator builds up standing waves, which is the condition for sound production and radiation. This process is particularly evident in large organ flue pipes (of 16′ and 32′ size), but can be observed also in smaller pipes (of 8′, see Figs. 30.17, 30.18 and [32.35]), in large duct flutes like the Slovak *Fujara* (of ≈ 160 cm tube length [32.41]), and even in large reed pipes where higher modes build up at their correct harmonic frequencies only after several hundred ms. In Fig. 32.6, the onset of the tone B_2 ($f_1 \sim 122.3$ Hz) recorded from a bassoon played with medium force (*mf*) is shown. One can see that mode no. 2 (corresponding to partial f_2 in the spectrum) is building up early while mode no. 1 needs more time and mode no. 3 as well as higher modes start with broad spectral lobes meaning a stable regime of vibration has not yet been established for these modes. Also, there is some noise in the transient signal in a frequency band above ≈ 1 kHz. The steady state of the tone as defined by clear spectral peaks

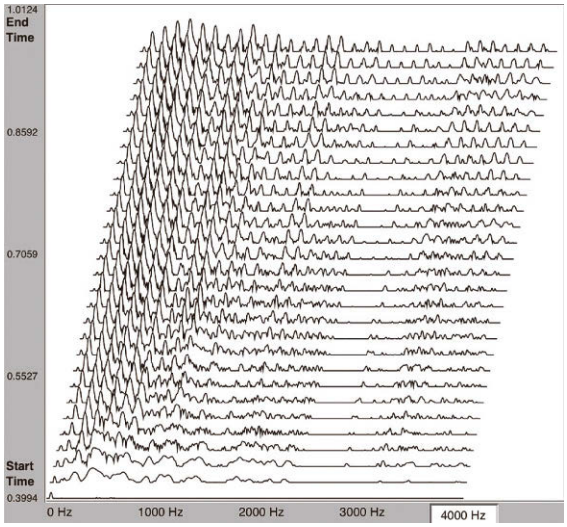


Fig. 32.6 Bassoon, onset of sound for a tone at $f_1 \sim$ 122.3 Hz; 29 spectra (FFT: 4096 pts, Hanning, hop ratio 0.2)

marking harmonics as well as by a small degree of spectral fluctuation is reached only after $\approx$ 300–400 ms. However, a clear pitch conveyed by a number of low harmonic partials as well as the period of the temporal envelope can be perceived much earlier (within $\approx$ 100 ms from onset).

The sounds from organ pipes analyzed in Sect. 30.2 and the bassoon sound (Fig. 32.6) demonstrate that, in particular in aerophones but also in plucked and bowed chordophones, a transient part precedes the steady state; the transient part results because the vibrating system (air column, string) has a certain input impedance and exhibits inertia. Since the input impedance for a cylindrical tube filled with air is

$$Z = \frac{p}{q} = \frac{p}{v\,A}\,,$$

where p is the pressure and $q = v\,A$ is the acoustical volume current (for a plane A and particle velocity v), pressure has to build up beyond a certain threshold before a stable vibration pattern with standing waves and radiation of a periodic sound signal is achieved. Consequently, the transient part contains noise in the signal in addition to the restricted number of harmonic partials resulting from such modes that respond to the excitation right from the onset. In plucked strings, as in the harpsichord, the transient comprises both a longitudinal wave preceding transversal motion as well as a short noisy segment resulting from the interaction of the plectrum with the string [32.36]. In strings excited with a small hammer, one can observe a short noisy precursor signal followed first by the longitudinal wave and then by transversal motion of the string (for measurements of a Stein-Conrad Hammerflügel from 1793 that has delicate small hammers, thin strings and a fast action see [32.42]). The precursor signals are quite short (usually, $t < 10$ ms) yet are of perceptual relevance (Sect. 32.2.3).

32.1.3 Organology, Electronics, and Timbre: Some Historical Facts

Humans evidently recognize and value different sound qualities given the immense number and diversity of musical instruments in use in various cultures and ethnic groups around the world. Exploration of the field known today as *sound color* and *timbre* was begun, on an empirical level, by musicians and craftsmen when making flute, reed, brass and string instruments. Historically, highly developed instruments such as the bronze lurs from Scandinavia [32.43] and bronze bells from China [32.44] testify instrument makers and musicians must have had a regard for sound quality already in antiquity. Timbre as related to certain materials played a significant role in Chinese tradition of instrument classification [32.45, p. 67 ff.]. Differences in sounds were also discussed in Greek writings on music theory (as in chapters of Aristoxenos' treatise on music, see [32.46]). In Roman times, several brass instruments, because of their powerful (as well as *horrible*) sound, were employed for military purposes while the first organs (the organum hydraulicum, said to have been invented, in Alexandria, by Ktsebios) appeared as instruments suited to fill a theater or arena with sound [32.47]. From the Middle Ages onward, mensuration of organ pipes perhaps led to a basic understanding of the interdependence of *pitch* and *sound color* in flue pipes. Musical instruments such as shawms, bagpipes and the hurdy-gurdy are mentioned quite frequently in medieval and Renaissance literature including remarks on sound properties. The development of bagpipes as well as the medieval bladder pipe not only provided players with a continuous sounding instrument but also with a *sonorous* sound quality (both chanter and drone(s) employed single or double reeds). Similarly, the hurdy-gurdy (which appeared in Europe around the 13th century [32.48]), offered a performance style based on continuous drone accompaniment plus distinct melodic pitch sequences as well as a specific sound (resulting from the interaction of the turning wheel with the strings, where the speed of the wheel and thereby the *attack* on the strings can be varied). Organ dispositions of the 14th century indicate that several instruments already offered a contrast of two sound concepts, namely diapason (*Prinzipal*) and organo pleno (*Grand orgue* [32.49, p. 10 ff.]). By the end of the 15th century,

the wall profile for the minor-third bell was discovered; a good example is the *Gloriosa*, Erfurt, cast by Geert (Gerhardus) de Wouw, in 1497. The minor third in this particular bell has ≈ 280 cent (the Pythagorean minor third has 294 cent), which suggests bell founders had an understanding of how to produce a certain spectrum by shaping the profile of the bell's wall. A detailed account of the instruments in use by about 1600 was provided by *Praetorius* [32.50]. The broad range of instruments (in particular reeds) developed in the Renaissance and the concept known as *Spaltklang* (split sound) realized in organ dispositions (as reported in [32.50]) as well as in ensembles demonstrates a keen sense of timbre. In modern times, *sound color* (German: *Klangfarbe*) became an essential feature in Western art music as reflected in orchestration (for a comprehensive survey, see [32.51]) growing increasingly more complex in works of the 19th and 20th century respectively. Composers demanded rare or newly invented instruments (such as the celesta employed by Tchaikovsky in *The Nutcracker* or the Heckelphone in Strauss' *Salomé*) as well as unusual ways of playing (as in many works of modern music after 1945) in order to have unique or even perplexing sounds at their disposal. It should be noted that the temporal and spectral characteristics of tones played on most orchestral instruments vary considerably in regard to dynamics (from *pp* to *ff*), to the effect that the radiation pattern and hence, the directivity of sound also changes a lot with dynamics (for a comprehensive survey of facts and data, see [32.52]). In this respect, the timbre is by no means constant (besides the variation of spectral energy distribution observed when instruments are played in different registers).

The advent of electroacoustic music brought technical sound sources (generators, oscillators) and devices to combine and modify sources (e.g., ring modulator, vocoder) into play [32.53–56]. The analogue (voltage-controlled) synthesizer and digital instruments that became available for common use in the 1970s and 1980s respectively, offered even more choices for creating complex sounds [32.56–58]. Digital synthesizers (like the Yamaha DX 7) as well as digital samplers (like the Akai S 1000 and the Emax II) in fact are music computers based on signal processing technology. The computer had been used for sound generation, by Max Mathews and other pioneers, for about two decades before digital audio became a standard in sound recording and music media (the CD was standardized as a digital format around 1979–1980 and introduced to the commercial user market in 1982). Special techniques like frequency modulation (FM) synthesis [32.59–61] allowed both to replicate the timbre of many existing instruments, among them idiophones like xylophones, gongs, etc. and to create sounds with complex, time-variant harmonic and inharmonic spectra, which were unprecedented. Such sound material led to compositions that transcend borders between the common categories of *pitch* and *timbre* (like *Stria* [32.62] by John Chowning, *Bossis* [32.63] and Sect. 32.2.4).

In addition to electronic and digital sound synthesis, all kinds of environmental and technical sounds were put to use in *musique concrète*, which, as one of its goals, considered aperiodic noises and periodic waveshapes as a continuum to be exploited for sound collage techniques [32.64, 65]. A far-reaching, more generalized concept of *sound* evolved in areas of electronic and computer music as well as in studio productions of pop and rock music when room acoustics, multichannel recording and reproduction and a host of audio effects (such as artificial delay and reverb, phasing, flanging, chorus; see articles by *Dutilleux* and *Zölzer* [32.66]) were integrated with orchestral and electronic instruments as well as with analogue and digital sound samples. In consideration of these developments, it became customary already in the 1960s to speak of, for example, *the sound of Mantovani* (multiple bowed strings deeply embedded in reverb) or *the wall of sound* produced by Phil Spector in pop music recordings, for which the recipe was to have many instruments played simultaneously in a rather small studio so that their sounds overlay and are hardly recognizable individually since Spector added, moreover, amounts of reverb and compressed the mix dynamically so that indeed a very dense *wall of sound* is audible in recordings like *Be My Baby* [32.67]. Later on, there was *psychedelic sound* in which time-axis manipulation of signals such as phasing and flanging as well as stereo panning effects figured prominently (e.g., Tomorrow: *Revolution*, 1967, Jimi Hendrix: *All Along the Watchtower*, 1968), and then *disco sound* (with huge concentration of spectral energy at the bottom end of the audible frequency range and typical patterns of percussion and bass in regard to meter and rhythm), etc. In all these *sounds*, timbre played an important if not decisive role. However, sound in this respect rather is a conglomerate of natural and artificial sound sources, effects, production and reproduction techniques while *timbre* traditionally (as in treatises on orchestration [32.12, 51]), has been assigned to single instruments or to the voice of certain male and female singers.

32.1.4 Research on *Timbre* and *Sound Color*: A Brief Retrospective

Though elements of acoustics can be traced in Greek antiquity (for example, the observation that the tension of a string determines whether the tone it pro-

duces sounds dull or sharp), a more systematic approach was pursued in the 16th and 17th century when empirical research was established along with mathematical treatment of problems. Knowledge concerning sound structure improved a lot when Beeckman and Mersenne understood the nature of harmonic vibration that led to the discovery of partials in strings (for which Sauveur gave a detailed description in the years 1700–1713 [32.68–70]). From Sauveur's published lectures, *Rameau* [32.71] saw that musical tones comprised a fundamental and its harmonics. By about 1800, it was clear to acousticians like Chladni that musical sound in general was a mixture of harmonic or inharmonic partials. The explanation *Chladni* [32.72, p. 241 ff.] gave was that elastic bodies can undergo very many vibrations at the same time, which would correspondingly activate many different parts of the inner ear without hampering each other. Thereby, sounds from different instruments could be perceived simultaneously. *Chladni* [32.72, Sec. 248] attributed different sound qualities to the different materials of elastic bodies consisting of organic or inorganic material (e.g., wood, brass, iron) and the microstructure of vibration inside such bodies as well as in the media (fluids, solids) through which sound propagates. *Opelt* [32.73, Sec. 7] held that the quality of sounds (e.g., strings, trumpet, flutes) or the *so-called sound color* (*Klangfarbe*) depends on different kinds or shapes of pulses reaching our ears. Opelt argued pulses and vibrations must be complex since, in a musical instrument, all parts vibrate; that is, strings vibrate coupled to resonance plates and air columns vibrate coupled to the tubes and bells of wind instruments. The resulting sound embedded in complex tones thereby is *an aggregate of several isochronous pulse sequences* having their origin in various parts of the instrument. Moreover, the mechanism of excitation (plucking or bowing a string, hammers in keyboards, etc.) would result in different *sound colors*.

When Helmholtz, in the 1850s, began his work on musical acoustics and psychoacoustics, he had tuning forks and resonators as well as sirens for sound analysis. Sets of precise tuning forks driven electromechanically for continuant tones provided kind of an early synthesizer (Rudolph Koenig, ten forks, $f_1 = 128\,\mathrm{Hz}$ [32.74, p. 329 ff.]) for complex sounds such as vowels. Koenig, himself an acoustician who cooperated closely with Helmholtz, also constructed a mechanical wave analyzer with a set of resonators. In the 19th century, some elementary tools for recording sound waves had become available [32.75] before Edison developed his improved model of the phonograph (issued in 1888) that was used in many investigations of sound. Among the objects of study was sound production in the human voice, and in particular the nature of vowels [32.74, p. 367 ff.]. In the second half of the 19th century, the quest for finding a specific resonance mechanism that could explain the production of vowels led to theories of *formants* (a term coined by the physiologist Ludimar Herrmann in the 1880s who contributed greatly to empirical sound research [32.34]). Sound research gathered further momentum when, after about 1920, continuous recording of sound on film labeled *phonophotography* [32.76, p. 10 ff.], [32.77] and analysis of the sound wave both in regard to periodicity and spectrum became widespread as lab techniques. Pitch had been calculated as fundamental frequencies from the periods of vibration (by $f = 1/T$) even before and spectral analysis had been done with the aid of tuned resonators, notably by Helmholtz. However, spectral analysis by means of filters [32.78] not only allowed identification of spectral components but also investigation of transient behavior. Using octave sieves where the impulse response is short due to the rather broad filter bandwidth one could see modes of vibration and corresponding spectral components building up over time before a stable (quasistationary) regime was reached [32.79, 80]. Still, *sound color* rather referred to the quasistationary regime of vibration and its corresponding spectrum showing the amplitudes of spectral components at their (harmonic or inharmonic) frequencies while transients were addressed as the short section at the onset where the sound wave often lacks clear periodicity (Sect. 32.1.2; for a survey of research on transients in nonpercussive instruments, see [32.31]).

Significant progress in musical sound research was made when an improved model of the analogue Sona-Graph became available for musicologists in the 1960s. The Sona-Graph, first issued in the late 1940s for research in phonetics (*visible speech*), allowed a spectral and temporal representation of sound in a quasi-three-dimensional (3-D) format (time as abscissa, frequency as ordinate, and energy of partials indicated by degrees of a grayscale). The Sonagraph Model II offered several bandpass filter settings and was used to explore sound structures in Western and non-Western musics [32.81] including characteristics of the singing voice [32.82]. In the 1980s, digital spectrum analyzers (such as the B&K 2032 model) allowed FFT-based spectral analysis of complex sounds [32.83]. Since the 1970s, a number of special codes for sound signal analysis had been developed including linear prediction, autoregressive models, and pitch tracking [32.33, 84]. From about 1990 on, powerful workstations suited for digital signal processing (DSP) became available. Software packages (like sndan, based on phase vocoder analysis/synthesis, introduced in 1993 [32.28]) allowed users to perform high-resolution sound analysis and synthesis/resynthesis. With the new tools (both hardware and

software) it was possible to study temporal and spectral structures of sounds recorded from Western and non-Western (e.g., Indonesian *gamelan*) instruments in great detail [32.85, 86]. At the same time, approaches to sound synthesis and resynthesis were refined further when wavelets, granular synthesis, digital waveguides and other techniques had been developed (there are collections of many relevant articles in [32.28, 66, 87, 88].

Summing up this paragraph, it seems obvious that exploration of musical timbre depended significantly on the tools available to researchers for sound analysis and synthesis at a certain time and place. There is a line of progress leading from mechanical to electrical devices and finally to computer-based algorithmic modeling of sound and sound-producing instruments and to ever more fine-grained analysis. Not only did acoustical and musical sound research benefit from computerized tools. As will be seen in the following section, also psychological research into timbre and sound color took new directions when computers and software for advanced statistics became available for many institutions. Meanwhile, toolboxes for sound research have been set up including audio descriptors applicable to musical timbre [32.89].

32.2 Sensation and Perception of *Timbre* and *Sound Color*

32.2.1 Classical Concepts of Tone Color or Sound Color

Helmholtz [32.90] ascertained that the pitch of a simple or complex harmonic tone depends on the period of vibration and that sound intensity depends on its amplitude. He attributed sound color to the microstructure within each period of vibration as well as to the fine structure of the resulting sound wave. He stated that each different sound shape calls for a distinct shape of vibration whereas several different waveshapes might bring about the same sound color. Different sound colors according to Helmholtz result from different patterns of harmonic partials added to the fundamental frequency (f_1) determining the pitch of the sound. Hence, differences in the sound color of complex sounds result from the number and amplitudes of partials above f_1 while phase differences between partials, *Helmholtz* [32.90, p. 194] declared, can be neglected. This view is in line with his resonance theory of hearing, which posits the inner ear would perform a Fourier analysis whereby a number of partials would become audible as constituents of a harmonic complex. From his observations *Helmholtz* [32.90, p. 97] argued that strong partials suited to activating a resonator would also be audible as an individual harmonic (*Oberton*), and that no *Oberton* was audible in the case where no response from a resonator had been observed. Though Helmholtz considered also sounds with inharmonic partials, his main concern was the harmonic type since most of the instruments assembled in an orchestra are chordophones and aerophones.

Stumpf [32.91, p. 520 ff.] stated that, from a phenomenological perspective, human subjects assign three basic attributes to sounds, namely pitch (*Höhe*), intensity (*Stärke*), and extension (*Größe*). While sensations of pitch and intensity can be directly traced to physical properties of sound (frequency and period as well as the amplitude of the sound wave reaching the ear), extension as implying some spatial interpretation is a more complex concept that can incorporate several sound features and sensory attributes. In his book on speech sounds, *Stumpf* [32.92, Chap. 15] also elaborates on his concept of instrumental sounds in detail. He distinguished *inner* and *outer* moments of sound color, where the latter depend mostly on temporal factors (transients, envelope, modulation) while the inner structure of sounds depends mostly on spectral composition and energy distribution [32.34]. Both taken together perhaps embrace what the term *timbre* seems to denote: an intricate combination and interplay of temporal, spectral, and dynamic features of sounds that, in normal sensation and even in an analytical mode of hearing, are often difficult to analyze and hard to separate from each other. *Stumpf* [32.91] argued that, in actual sensation and perception, even basic attributes are not completely independent since the pitch of pure tones can vary to some extent with intensity, and both combined give rise to differences not only in *tone height* but also in brightness, density, and volume. *Köhler* [32.93–96] and some other researchers pointed to the similarity between pure tones played from low to high frequencies and the sequence of vowels *u-o-a-e-i*, resulting in an attribute often labeled *vowel quality* (also termed *vocality*).

The phenomenal description of tonal attributes and their interrelations had been addressed, from an empirical descriptive approach including auditory tests with musically trained and untrained subjects, by *Stumpf* [32.91, 92, 97–99] and by several of his coworkers [32.93–96, 100] as well as by other researchers. For example, *Rich* [32.101] had proposed the attribute of *volume* to mark the spatial extension or diffusion of low tones against high ones imagined by listeners in addi-

tion to other such attributes like brightness or density. *Stevens* [32.102] found that even pure tones evoke sensations of *volume* (*bigness*, *spread*, [32.103]), and that this attribute relates to both the frequency and the intensity of tones. Since pitch to a small extent [32.104] and loudness attributed to pure tones also depend on both frequency and intensity, *Stevens* [32.102] warned *that no psychological dimension need be a simple correlate of a single dimension of the stimulus*. In fact, several attributes covary positively or negatively with frequency and intensity; an increase in frequency from low to high brings about sensations that covary positively with frequency, namely pitch (if taken as *tone height*), brightness, and density. Brightness increases also when intensity is raised within certain limits, and the same holds true for density as a phenomenal attribute of sound. Further, increases in intensity for pure and complex tones in the treble frequency range ($\approx 6-10$ kHz) can turn sensation of brightness into unpleasant sharpness. While tone height, density and brightness increase with frequency, the volume of tones seems to be larger at low and smaller at high frequencies (provided constant sound pressure level (SPL)). *Lichte* [32.105] confirmed that, besides pitch and loudness, complex tones have *at least three attributes. These are brightness, roughness, and one tentatively labeled fullness* (which corresponds to *volume* in other studies).

Stimulus parameters and attributes of sensation that had been described in a range of studies [32.92, 100, 103, 107–109] were given a detailed interpretation by *Albersheim* [32.106]. Ordering tonal properties that have objective correlates in physical parameters as well as tonal attributes (these are phenomenal descriptions of sensations) in regard to their mutual relations, he derived a scheme for pure tones and another scheme for complex tones of which slightly adapted versions (to account for differences between the original German terms and English translations) are shown in Fig. 32.7a,b.

Likewise, *Albersheim* [32.106, p. 268] condensed his discussion of the stimulus parameters and sensory attributes of complex tones into the scheme shown in Fig. 32.7b.

The term mixed color denotes the phenomenal quality of a sound resulting from the composite effect of fundamental frequency and spectral energy distribution. Different spectral patterns varying in the number and strength of partials determine the sensation of brightness as well as the similarity such complex sounds may have with vowels in speech and singing, on the one hand, and timbres known from musical instruments, on the other. In addition, the phenomenal quality of a complex sound can change when it is shifted up and down a musical scale. The term *mixed color* expresses

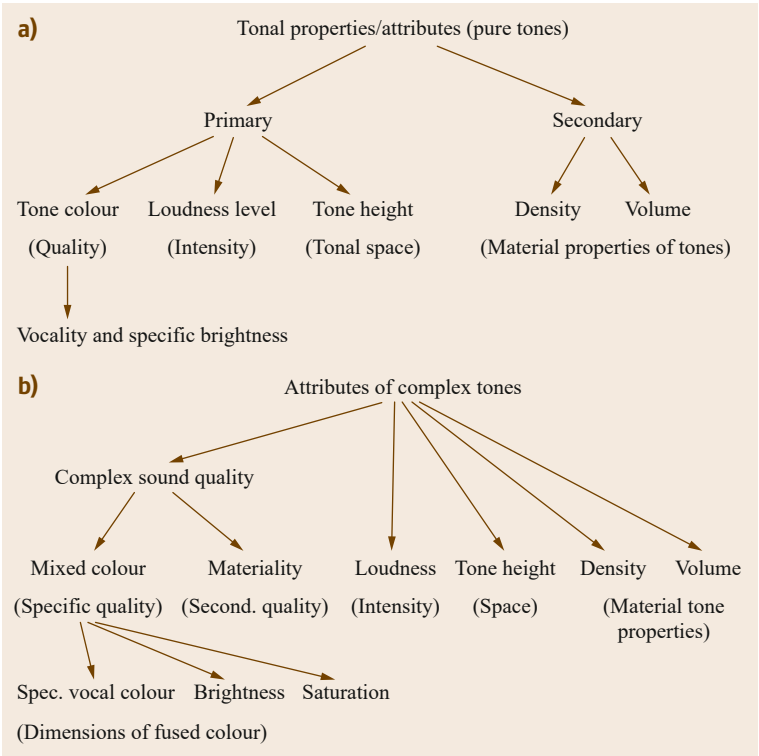


Fig. 32.7 **(a)** Tonal properties and attributes for pure tones (after [32.106]) **(b)** Properties and attributes of complex tones (after [32.106])

the combined effect of all these factors. Materiality (German: *Stofflichkeit*) denotes material properties of a sound one can *touch* and *feel* (analogous to haptic sensation), for instance, roughness or smoothness of sounds. *Albersheim* [32.106] devoted a significant part of his monograph to phenomena and categories known from the study of phonetics and voice quality in regard to their applicability to musical sound, addressing aspects like *specific vocal color* of complex harmonic tones. Vowels had been covered by *Köhler* [32.93–96, 100] and given systematic treatment by *Stumpf* [32.92] on the basis of many experimental findings.

In contrast to the elaborate descriptions of tonal attributes provided by Stumpf, Hornbostel, and Albersheim, the cognitive psychologist *Ebbinghaus* [32.110, p. 306] gave a neutral definition of sound color when he said the *sound color* (*Klangfarbe*) of tones is understood as that which distinguishes them in sensation, at identical pitch and intensity, when produced from different instruments or voices. This statement implies that there are distinct sound properties giving rise to (at least three) basic sensations: pitch, intensity, and sound color. However, the nature and perceptual *content* of sound color were left out of this definition which, much later, resurfaced in nearly identical shape as that of *timbre*, issued by the American Standard Association ([. . .] *the attribute of sensation in terms of which a listener can judge that two sounds having the same loudness and pitch are dissimilar* [32.111]).

32.2.2 Modern Approaches: From the *Objet Sonore* to Multidimensional Scaling

As is evident from the preceding sections, approaches to the classification of sounds in regard to sound color or timbre originally started from fundamentals of vibration, sound production and radiation as realized in certain instruments (including the voice). As far as sensation and perception is concerned, the basic concept was that of psychophysics where one seeks for a physical input parameter causing an output in sensation that is measurable (and possibly scalable in some unit, Sects. 30.1.2–30.1.4). This view was gradually but steadily changed when, in the descriptive and analytical studies published by Stumpf, Rich, Hornbostel, etc. the experience of subjects in perceiving sound played an increasingly greater role. Though physical and physiological facts known from measurements and observations were still taken into account, the focus in music psychology was on tonal *attributes* relevant for perception. The phenomenological paradigm reflecting the experience of musically trained subjects is clearly evident in the monograph *Albersheim* [32.106] devoted to attributes of tone and sound.

In the following, a number of approaches to the study of timbre and sound color will be discussed in regard to developments in music and media technology as well as in empirical research methodology. For obvious reasons (given the large number of publications on timbre, and on *sound* and perceptual aspects of sound in general), the survey must be selective.

Developments in Audio Technology

Technical production and reproduction of music as sound began after the phonograph and the gramophone had been introduced to the public [32.112]. Radio transmission followed in the 1920s when electric amplification and recording on the basis of vacuum tube technology was developed, and solid-state technology (the transistor was invented in 1947/48) came into use during the 1950s. By about 1930, the *Trautonium* (a kind of early synthesizer) had been invented, followed soon after by its polyphonic expansion, the so-called Mixtur-Trautonium; these instruments were used by composers of modern music (P. Hindemith, H. Genzmer). Also in the 1930s, the Hammond organ became available in several models (making use of rotating tone wheels and electromagnetic pickups for sound generation). Shortly before and during World War II, the modern tape recorder was developed (even in stereo, by AEG). The vinyl LP (12″, 33 1/3 rpm) was issued in 1949. Evolving technology offered new possibilities to creative artists engaged in what became *electronic music* [32.53]. Institutions like the Studio for Electronic Music at Cologne (part of the public radio station, NWDR, later WDR, and home turf of composers Herbert Eimert and Karl-Heinz Stockhausen) offered technical facilities including a range of sound generators, filters, ring modulators, and tape recorders as well as electronic keyboard instruments (for technical information and musical aspects, see [32.113]). *Moles* [32.114, Chap. IV] elaborated on sound according to physical (time, frequency, SPL) and psychoacoustic parameters (envelope, spectrum, level and dynamics, pitch, timbre), which can be used for both analytic description of existing *sonic objects* as well as for the creation of new ones. His main idea was that music as is performed, recorded on tape or other medium, and perceived by listeners consists of *objets sonores* that can be distinguished from one another, and can be decomposed into cells in a frequency/level plane representing the area defined by lower and upper limits of human hearing. *Moles* [32.114, p. 118–119] considered such cells (he uses terms like *cellules* as well as *quanta sonores*) as carriers of information, thus relating to concepts of communication developed in a more formal

approach by *Gabor* [32.115] and by *Shannon* [32.116]. Moles also offered rules of how to work with sonic objects. The acoustic and psychoacoustic description of sonic objects met the scientific spirit encountered in electronic music circles. The tape recorder became instrumental in collage and montage techniques employed not only in electronic music but also in *musique concrète* that made use of a broad range of natural and technical environmental sounds [32.64]. The idea to expand sound material from tones and chords as well as complex sonorities into *musical noise* had been propagated by Italian *Futuristi* [32.117] and had been pursued by some composers in Europe and overseas.

Schaeffer's Typology of Sonic Objects

An attempt to deal with the huge range of sonic and musical objects serving as material for contemporary music was the comprehensive *solfège des objets musicaux* prepared by *Pierre Schaeffer* [32.118, pp. 387–597]. *Schaeffer* condensed his considerations into a scheme [32.118, pp. 584–587], which is formally organized like a matrix (*m* rows, *n* columns). The rows contain what he calls criteria of musical perception (critères de perception musicale); the criteria are masse, dynamique, timbre harmonique, and profil mélodique, profil de masse, grain, allure. These terms refer to volume (in the sense of Stumpf, Rich, Stevens), dynamics, and spectral as well as tonal structure in sonorities. Since Schaeffer's scheme is not confined to single sounds, and in fact tries to establish a typology of musical objects, he includes basic types of melodic contour (like the *podatus* and *torculus* known from Gregorian chant and neume notation). Grain (on the level of sounds) refers to the *surface* (which can be rough or smooth), and allure means the envelope, which can be regular (with or without vibrato) or irregular, etc. The columns comprise qualifications (types, classes, genres) and evaluations according to *species* (espèce, which also can designate a sort or a kind) in regard to the *height* (hauteur), intensity, and duration of sonic objects. To be sure, the typology offered by Schaeffer was a bold attempt which, however, is based on personal experience and description rather than on systematic treatment backed by empirical data. As a typology, it serves to order phenomena according to certain aspects even if it is not consistent in every respect (the matrix has some empty cells). Like in the phenomenological description of tonal attributes [32.106], the perspective is that of the appreciative subject perceiving sound and music.

Semantic Attributes of Timbre

Though subjects experience sounds as sensations characterized by certain qualities and intensities, perception can involve verbalization when judgment based on sensory data includes an act of predication [32.119]. Since sounds and their timbres allow for verbal description (and, moreover, can convey a musical or extramusical meaning), people speak of a *rumbling thunder*, *gurgling water*, *whistling wind* as well as of a *whining violin* or a *blaring trumpet*. Most languages contain very many adjectives used to characterize certain sounds or their timbres, and one can find a host of such adjectives in reviews of concerts and recordings as well as in other publications on music. In psychology, sets of adjectives were used to characterize expressive qualities of musical pieces or phrases in order to disclose their affective mood and their *meaning*, in regard to listeners [32.120].

The study of semantic meanings that things or persons or ideas have for various people was pursued by *Charles Osgood* et al. with a methodology known as semantic differential [32.121] and also as *polarity profile*, in areas of psychology. The goal is to measure attitudes and preferences of people as they value or dislike certain things, persons, ideas, etc. Osgood et al. conceived of a *semantic space* in which various concepts (this term was used instead of *stimuli*) could be placed according to the judgments subjects give on lists of adjectives (or nouns) ordered so as to form bipolar pairs (like soft–hard, good–bad, fast–slow). In between such polar adjectives, a scale is inserted for which *Osgood* et al. [32.121, p. 85] proposed seven alternatives. These can be expressed by numbers or by verbal qualifications suited to indicating a scale. The technicalities of scale construction are in fact as important for the outcome and the reliability and validity of experiments as is the selection and the arrangement of adjectives (some critical issues are discussed in [32.122–124] and [32.125, p. 73 ff.]).

A semantic differential or polarity profile usually comprises from about 30 to 70 pairs of adjectives, which are the variables used to characterize a number of items (*concepts* in Osgood's terminology), by a sample of subjects. The resulting data (blocks of variables × items × subjects) is subjected to the calculation of descriptive statistics (means, variances, etc.) and of intercorrelation matrices, on which in many studies factor analysis (FA) with respect to PCA (principal component analysis, usually with varimax rotation) has been performed. Without going into details of FA (besides PCA, there are several other methods and models in use), it should be noted that the methodology involves a load of vector and matrix algebra as well as the representation of the variables and the factors calculated from the variables in a *k*-dimensional vector space [32.126]. To be stable as topological constructs and reliable and valid in regard to interpretation, fac-

tor models must conform to geometrical axioms and principles. *Osgood* et al. [32.121, p. 91] suggested to conceive of distances between *concepts* in the semantic space in terms of linear geometrical distances. To calculate these distances without violation of geometrical axioms (and minimizing the risk of artifacts), the raw data should be interval or ratio scaled. However, this is a condition rarely fulfilled in studies based on semantic differentials. Moreover, these seem to comprise two different kinds of scales as the adjectives selected for the polarity profile in many studies are either denotative (descriptive in regard to features and attributes of items) or connotative (associative in regard to features of items); some adjectives are in between these categories. Consider, for example, sounds recorded from various orchestral instruments (strings, woodwinds, brass, percussion) that differ in temporal envelope and spectral structure (Sects. 30.2 and 32.1.2). On the denotative level, one could offer adjectives such as soft–loud, dull–bright, smooth–rough, transparent–dense. Further, there are adjectives such as *nasal* that are widely used but may refer to more than one acoustic property [32.127]. In addition, there are adjectives that may denote timbre qualities more indirectly. For instance, the sound of a trumpet appears *shining* to many listeners; the objective basis for this is spectral energy distribution and centroid. Then there are adjectives that express connotative rather than denotative meanings associated with certain timbres like *doleful*, *exciting*, *gloomy*, etc. The problem is that the scales used for denotative attributes can be regarded as rating scales (subjects try to estimate a certain quality and/or intensity on the basis of a sensation) while the connotative adjectives rather call for an emotional appraisal or associative guess (consider, for instance, a pair like *dreamy–awake* in regard to the violin sound of Isaac Stern playing the opening measures of the Adagio in Bruch's violin concerto).

The semantic differential (often coupled with FA) has been used in the study of musical and other sounds [32.128–132]. In *Osgood* et al. [32.121, p. 36 ff.], the *standard* solution obtained from FA consists of three factors, the first of which was interpreted as evaluative, the second as a *potency variable*, and the third as *activity*. Such a solution restricted to three factors seems plausible if the factors allow for a rather wide interpretation and thus can embrace a larger number of variables (in this case, pairs of adjectives). It should be noted that *Wundt* [32.133] had proposed a three-componential model of emotions based on three polar pairs (pleasure–displeasure, excitement–inhibition, tension–relaxation) from which a number of additional emotions were derived as combinations (e.g., joy is derived from excitement, pleasure, and tension). In a factor model consisting of three independent, that is, orthogonal factors, these fit into a three-dimensional space so that the items can be conceived as *points* (relative to coordinates x, y, z in Euclidean space) and the distances between items (as well as between subjects) calculated as Euclidean distances where the linear distance between pairs of points ($a = x_i, y_i, z_i$; $b = x_j, y_j, z_j$) is interpreted as a measure of phenomenal similarity or dissimilarity respectively. Finally, three factors in many empirical studies suffice to explain a substantial percentage of the variance in the data. However, the factors derived from FA again are vectors that do not always lend themselves easily to factual or perceptual interpretation in regard to the items and/or subjects. Also, the interpretation calls for some *label* that will be attached to a factor. *Rahls* [32.129] found for 20 sounds and 47 pairs of adjectives a four-factor solution of which the first three were interpretable: the first as an *evaluative* factor, the second as one that involves *activity* and the third viewed as *potency* (see above). In *Jost*'s [32.130] study of clarinet sounds, one factor pointed to *volume* as an attribute, and a weak factor indicated the specifics of the clarinet timbre.

In contrast, *von Bismarck* [32.131] found for his sounds (sine tones, complex tones, noise) a strong factor accounting for 44% of the variance that relates to spectral energy distribution and presence of energy in higher frequency bands. This factor, which quite clearly reflects a sensory attribute [32.134], was labeled *sharpness* (German: *Schärfe*). In an experiment taking up some of Bismarck's pairs of polar adjectives as *differentials* but using dyads played from wind instruments, *Kendall* and *Carterette* [32.132, p. 455] found that *sharp is not a good discriminator across wind instrument dyads* since *sharp* in English refers to *pitch* rather than timbre. The problem encountered with such terms as *Schärfe* ≠ *sharpness* of course is one of semantics. Many of the terms used in description of sounds, even if denotative in essence, are taken from other spheres such as optics (brightness) or haptics (roughness, smoothness). Certain attributes are intermodal in regard to sensation as is the case with brightness [32.135] and apparently so with roughness (which is scalable both as an auditory attribute of sound and as tactual experience; *Zwicker* and *Fastl* [32.136, Chap. 11], *Stevens* and *Harris* [32.137]). Also *Schärfe*, which one could tentatively translate as *stridency* can be addressed as an intermodal attribute. The German *Schärfe* refers to the condition of a blade of a sword or knife that has been sharpened; the word also denotes the sharp outer edge of a swinging or carillon bell profile. Thus, the original meaning is in the tactual and haptic area of sensation. The term can be applied to sounds sensed as glaring, strident, penetrating and perhaps even *hissing* if they contain a strong proportion of energy in higher parts of the audible fre-

quency range. Such sounds can be easily generated in harmonic complexes when amplitudes of partials increase in proportion with harmonic number [32.138]. Another method simply is bandpass or high-pass filtering of harmonic or inharmonic complexes or noise bands so that energy is concentrated in the frequency range corresponding to higher critical bands (CBs) of the auditory system.

The problems involved in verbalizing sound and musical phenomena in general are complex and difficult to solve. In certain ways (reflected in a number of essays by *Charles Seeger* on fundamentals of musicology [32.139]), it would be recommendable, though perhaps not always practical, to communicate in the medium of music about things musical instead of using speech. From the viewpoint of empirical methodology, it seems wise to reduce semantic ambiguity in verbal descriptions of sound as far as possible, which can be done by selecting such adjectives and nouns that denote certain sound characteristics rather than using connotative adjectives (where subjects in a sample often have ideas as to their meanings that differ widely). Also, verbal descriptors should be checked in experiments for consistency relative to various sets of stimuli as well as for their strength of differentiation between items (one of the issues with semantic differentials is the lack of standardization [32.122]).

In order to avoid pitfalls that can happen with the methodology of semantic differential and FA interpretations, the use of adjectives combined with unipolar scales and a robust method of data analysis such as hierarchical cluster analysis seems advisable. This approach was chosen by *Thies* [32.140] in an attempt to find fundamental categories for a descriptive classification of sounds that could bring Schaeffer's more intuitive scheme to a systematic and formal typology. Starting from the German vocabulary which, in regard to sound and its attributes, offers some 1600 descriptive and connotative words, he selected 51 *general* classifiers (e.g., loud, soft, rough, smooth, high, low) and 382 more *specific* (like nasal, creaky, buzzing, pulsating, whispering). In experiments offering musical and environmental sounds as stimuli, subjects were asked to judge which of the classifiers applied to certain sounds, and if so, to what extent. Cluster analysis found groups of adjectives that represent basic descriptive sound categories. From these groups, pairs of adjectives (like tonal–noisy, dark–bright, soft–hard) representing fundamental spectral and temporal features and sensory attributes (tonalness, brightness, loudness, etc.) were considered as fundamental classifiers to be used in a first-level analysis and classification while, on the next level, more specific classifiers are used for a fine-grained classification and typology.

The problem of semantic meanings adjectives have with respect to properties of sounds and attributes of timbre may aggravate if viewed from an inter-language perspective. Inasmuch as languages reflect cultural norms and experiences shared by ethnic and language groups, it is likely that cultural differences manifest themselves in different concepts and terminology even though acoustic properties of sounds and also parameters of auditory perception are identical among such groups. In a comparative study with subjects having either Greek or English as their native tongue, the convergence of descriptive adjectives was tested with the verbal attribute magnitude estimation (VAME) method [32.123, 141, 142] which avoids semantic differentials by putting adjectives against their negation (e.g., dull–not dull, sharp–not sharp). This comparative study [32.143] employed various methods of data analysis (cluster analysis, FA, correlational techniques) and also feature extraction from the stimulus sounds (23 musical instrument tones). The data analysis resulted in three common dimensions labeled luminance, texture, and mass, where luminance refers to attributes like brightness and sharpness (to include, however, depth/thickness in the Greek sample) while texture was interpreted as related also to spectral energy distribution, and mass possibly referring to spectral density and flux. The results demonstrate that a set of sound stimuli presented to subjects from different culture and language groups may elicit responses that converge to a certain extent due to acoustic stimulus features while there are also differences in conceptualized attributes derived from descriptive adjectives. In a Swedish study on the timbre of the steady state of alto sax sounds also employing adjectives along with VAME, *Nykänen* et al. [32.144] found a combination of rough and sharp (seemingly included in the Swedish adjective rå ≈ raw, bleak), soft, warm as well as a vowel quality o-like to describe the sounds best. There was a correspondence to psychoacoustic parameters in that the attribute *rough* could be predicted from the model of roughness used by *Aures* [32.145, 146] combined with the model of sharpness proposed by *von Bismarck* [32.131, 134].

Similarity Ratings of Sounds

Another approach to timbre research is that of similarity ratings. Detection of similarity in objects conveyed as sensory input is an experience basic to humans and other species. There are several theories of similarity some of which emphasize extraction of features from sensory input that are used for comparison while others underpin the cognitive evaluation process underlying similarity judgments [32.147, 148]. In regard to the problem of verbalization addressed in the previous paragraph, one might assume that the recognition

of similarities driven by sensory input works without verbalization while an analytic cognitive evaluation of input rather would involve such. Assuming a hierarchical model that leads from sensation to perception and apperception (Sect. 30.1.2), different levels of processing would likely be reflected in shades of awareness of such processes.

In the field of sound and music perception, *Stumpf* [32.149, Sects. 6,7] especially argued that different degrees of similarity as perceived by subjects can be expressed as differences in distance, which implies similarity is scalable and differences in similarity can be translated into some distance measure. Scaling of similarity has been a paradigm for a long time, for which a standard methodology usable in psychophysics and other areas was developed [32.150, 151]. Since phenomenal similarities between a set of objects often rest on several or even many properties, the geometric concept according to which perceptual *similarity* ≡ *proximity* has led to a multidimensional perspective where objects can be represented as points in a k-dimensional space. The interpoint distances relative to a number of dimensions then express the *similarity* or *dissimilarity* between n objects on m dimensions ($n \gg m$). In addition, one can calculate and geometrically represent similarity judgments of different subjects in a sample to compare their perceptions. The methodology to carry out such empirical studies is multidimensional scaling (MDS), also termed multidimensional similarity structure analysis [32.152, 153]. There are some ideas and concepts shared by both FA and MDS, namely the idea that a subject's responses to complex stimuli comprising several or many variables can be explained and/or predicted from a small number of factors or dimensions as well as the concept to represent relations between high-dimensional (multivariate) stimuli in a rather low-dimensional geometrical space. This space, spanned as factor configuration or MDS model derived from empirical data in several steps, is assumed to reflect perceptual and also cognitive evaluations of subjects. It should be noted that the metric-dimensional approach to *similarity* perception has been challenged on methodological and factual grounds [32.147] and has also been defended as a basis valid for the study of similarity as well as recognition and identification processes [32.154–156]. Though MDS (and, likewise, FA) can be used for data reduction from a larger number of variables presented to subjects in experiment to a few metavariables labeled *factor* or *dimension*, its main function is to help shape hypotheses concerning the dimensionality of complex structures. Like in FA, MDS models can be derived in a precise manner (for Euclidean space and distance functions) if the input consists of interval- or ratio-scaled data. However, similarity judgments are ordinal as subjects find two objects (say, two sounds recorded from any two instruments at the same pitch and presented at the same loudness) are *not similar*, *somewhat similar*, *similar*, *very similar*, *highly similar* or even perceptually *identical*. To allow for a so-called nonmetric MDS, one can assign numbers to such degrees and turn them into proximities from which, in an iterative process (minimizing errors and adjusting interpoint distances between objects or other items), the *final* distances in a k-dimensional space are calculated. The distances of n objects relative to m dimensions are regarded to reflect the perceptions and judgmental decisions of the subjects. The type of metric (Euclidean, city block, Minkowski [32.152]) that is chosen for the model expresses different cognitive decision strategies. If the similarity judgment is decomposable into several more or less separate evaluations of similarity along dimensions $(a, b, c, \ldots, k)$, a so-called city-block metric reflects the additive process. In case a direct and overall estimate of the similarity between objects is performed, a Euclidean distance model sums the differences on the contributing dimensions. When dimensions have different salience for individual subjects, a weighted Euclidean distance model seems apt (though this technique requires certain precautions). A Minkowski r-metric can be chosen if, in a decision process with respect to similarity, one stimulus feature is predominant so that the dimension to which it relates seems supreme against all others. The goodness-of-fit that various models yield with respect to their metric, number of dimensions, error score, as well as the coefficient of determination in relation to the dataset, can be used in a cautious interpretation of the evaluation and decision processes involved in the perception of phenomenal similarity.

Since about 1970, computer software for metrical and nonmetrical MDS has been available, which has spurred many experiments also in the field of sound and music perception [32.157, 158]. Some of the now *classical* experiments on timbre were performed at Stanford [32.159–161]. Grey worked with 16 synthesized sounds that emulated those of orchestral instruments since one task of his research project was to explore computer-based analysis-synthesis techniques. Data from an experiment with 20 musically sophisticated subjects judging the similarity of the 16 sounds were subjected to nonmetric MDS, which did yield three dimensions interpretable in terms of acoustic properties of sounds. The first dimension quite obviously relates to spectral energy distribution and spectral envelope; the second was interpreted as relating

> *to the form of the onset-offset patterns of tones, especially with respect to the presence of synchronic-*

ity in the collective attacks and decays of upper harmonics. [32.159, p. 61]

The third dimension was also viewed as temporal and related to the energy distribution in the attack of tones.

Studies of sound and timbre in most cases have confined their MDS analyses to three or even two dimensions [32.132, 159–168], [32.169, Chap. 6] one of which seems almost notoriously related to spectral energy distribution, centroid and brightness [32.170]. The second often has been interpreted as related to onset characteristics (hard or soft attack, percussive or rather mellow sound) or to the temporal envelope in total. A third dimension sometimes has been viewed in regard to spectral flux which, in most sounds, is considerable at the onset of sound because of the transient portion and becomes small once the steady state is reached. Spectral flux (SF) can be calculated from digitized audio signals where SF expresses the rate of change in spectral composition from one frame of analysis to the next [32.171]. Some studies could not confirm the interpretation of a third dimension as expressing SF [32.158, 167] though in fact most natural sounds exhibit some or even considerable variation in both spectral frequencies and amplitudes over time. In case composite waveshapes are used as stimuli, one dimension can be interpreted as spectral energy distribution (sparse harmonics–rich harmonics) and another in terms of the *surface quality* of sounds; since a sawtooth with many harmonics in zero phase has steep slopes in every period, the sound appears comparatively *rough* while a triangular wave appears quite *smooth* [32.168].

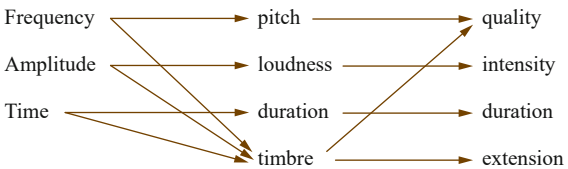
Incorporating a third dimension in a MDS model can be useful to account somewhat better for the variance in the similarity data; however, one might be confronted with a certain trade-off since, with a third dimension in a MDS model, as with a third factor in FA, one often may explain a higher percentage of the variance, on the one hand, while the interpretation of a third dimension or factor can be arduous, on the other. Not only does each factor or dimension need to be given a *label* (the verbalization problem is encountered again, if on another level), but its interpretation must also account for the factual *content* and perceptual effect of stimuli weighted in the light of the *k*-dimensional model. In general, interpretability seems to diminish with the number of factors or dimensions respectively. This can be possibly explained in regard to conditions under which similarity judgments are mostly made in real life, on the one hand, and processes of categorization, on the other. Concerning conditions, one has to take the huge amount of environmental information arriving at our senses as possible input into account. From an evolutionary perspective, *survival* makes very fast recognition of objects imperative. In a first and probably pre-attentive step, recognition involves some overall and provisional categorization in which comparison of a possibly new object to previously learned *prototypes*, *schemata* (or whatever the template is called) is performed. For an instantaneous comparison which, with respect to *survival* behavior, may trigger further sensory processes as well as motor responses, restriction to a few salient features seems pertinent. Adopting either a feature detection approach or that of dimensional metric as in MDS, reduction of the processing load is inevitable in fast estimates. The goal can be achieved successfully through exclusion of irrelevant as well as dimensional reduction of usable sensory input [32.172]. Even in many instances where the time factor is less critical, overall judgments of similarity apparently still seem restricted to a few salient features or a small number of dimensions respectively. Apparently, such suffice for a rough perceptual ordering of things into categories; listeners hearing music in a concert or from the radio can perform at least a tentative classification of sounds representing, for example, strings, brass, and percussion instruments. This can be done with reference to very few acoustical features and perceptual attributes, a fact corroborated from computer-based classification of musical sounds with respect to sources [32.173]. Many of the two-dimensional MDS models of *timbre similarity* in fact relate to the temporal and the spectral envelope as fundamental components of complex sounds; the two most common dimensions *centroid/brightness* and *onset/envelope* (see above) apparently suffice to categorize a host of sounds according to their timbre, at least in a gross mode. A closer inspection then may reveal further details needed for a finer classification. In this respect, timbre can be viewed as an emergent perceptual quality [32.174]. Of course, learning plays a role in such processes as recognition and identification of timbres where trained musicians usually perform faster and more reliably than nonmusicians. For example, it needs no expertise to assign a certain impulsive sound one hears to a class of instruments (say, membranophones) while one must have some experience to judge what the type of drum from which the sound comes could be (size, single- or double-headed, beaten with hand or stick), and it is certainly demanding to decide whether a drum sound correctly identified as that of a frame-drum belongs to a Moroccan *bendir* or to an Irish *bodhran*. However, in experiments involving synthesized FM sounds (TX 802) most of which emulated familiar instruments, differences in performance between subjects in three subclasses of musical expertise (professionals, amateurs, nonmusicians) were not just as great as one might have expected [32.167].

32.2.3 Acoustical Features and Perceptual Attributes of Timbre

Schouten [32.175] saw five acoustical parameters defining timbre:

1. Tonalness versus noise
2. Spectral envelope
3. Time (or temporal) envelope
4. Change in spectral envelope or in pitch
5. Onset characteristics, which he labeled *acoustic prefix* of sounds.

These in fact are interrelated in a number of ways, as are the four fundamental qualities of pitch, loudness, duration, and timbre that are constitutive for auditory perception. It is evident that one cannot perceive a *pitch* from a tone unless it does have a duration and sound intensity. Likewise, loudness as based on intensity and frequency [32.176] and variable also with the duration of stimuli presented to subjects calls for a sound that must have some frequency content with nonzero duration and amplitude(s). In this respect, parameters are not totally separable even though one can study the relative weight they can have in sensation and perception of complex sounds. The idea that parameters might be separable and decomposable into *dimensions* perhaps owes to classical concepts of psychophysics (Sect. 30.1.4) where sensations were studied with respect to quality, intensity, duration, extension, etc. Taking acoustical sound parameters, on the one hand, and perceptual dimensions, on the other, one can set up the following scheme:



Of course, pitch can be explained also in the time domain (Sect. 31.4), and loudness depends not only on physical intensity of the sound but to some degree on spectral energy distribution as well as on presentation time and temporal integration (Chap. 33). Further, timbres can be perceived as distinct qualities in case they offer salient features.

The view according to which pitch is separable, in principle, from timbre is not identical with that advanced by *Helmholtz* [32.90] on separability of pitch and sound color (Sect. 32.2.1). Taking the Fourier-based resonator model of Helmholtz, a harmonic complex simply is decomposed, on the basilar membrane (BM) level, into partials along a log frequency axis where the lowest partial (f_1) accounts for the sensation of pitch, and the remainder of the spectrum for *sound color*. The concept indeed implies that a sound, well-defined as to its pitch through f_1 (as well as its inverse, the period $T = 1/f$), gains some additional *coloring* from the spectrum above f_1. In this respect, sound color is a more or less stable quality corresponding to the steady state of a sound [32.177]. Further, sound color can be viewed in regard to *extension* where the number and strength of partials define spectral width, density, and centroid (for a given f_1 of, say, 110 Hz, the centroid shifts upwards in frequency with the number of additional partials). This effect is easily demonstrable with an analogue synth where a tone (say, A_2), when played with a sawtooth wave, gains in *color* as well as in brightness from sweeping the cutoff frequency of a low-pass filter toward higher frequencies. In early works on orchestration [32.3] the term *timbre* also referred to a sound quality largely determined by spectral structure while the temporal aspects as yet were of little concern. Since the 1920s and 1930s, when onset characteristics and temporal fluctuations of sounds could be studied with electroacoustic equipment at hand, transients in particular in aerophones and chordophones have been a topic of research (for a survey, see [32.31]). With computerized sound analysis, transients were investigated as dynamic time-frequency structures from which information relevant for auditory perception can be drawn [32.178]. In the following, the parameters discussed by *Schouten* [32.175] will be examined further in the light of research data.

Transients and Onset

The transient part of a sound reflects a vibrating system immediately after excitation, which can be effected by a single impulse as in many idiophones (e.g., xylophones, gongs, bells, cymbals) and membranophones as well as in plucked strings, or by a sequence of pulses as in aerophones (Sect. 32.1.2) and in bowed chordophones where excitation continues with energy supply to the generator. The transient portion of a sound can be defined as that from the absolute onset of vibration (the time point or sample where the amplitude is nonzero) up to a point where vibration becomes either periodic with small fluctuations (as in aerophones and chordophones) or where the peak amplitude is reached and the decay of the envelope begins (as in idiophones and membranophones excited by a single impulse). The transient portion of a sound is the most *interesting* part in terms of information (see [32.25] for concepts describing information structure of sounds) since information depends on entropy and the rate of change per time unit. *Shannon* [32.116] gives a formal proof that white noise has the maximum possible entropy. In comparison to the almost periodic regime of vibration in the steady state, the onset often lacks clear periodicity and

can even appear *chaotic* (as indicated by the limit cycle for the time series of a vibration or sound in a phase space). The onset of many sounds recorded from natural instruments includes noise in the transient portion; a well-known phenomenon observed in organ flue pipes is the *spitting* response to air pulse sequences before a standing wave is established (Sect. 30.2 and examples in [32.179, Chap. 5]).

For the listener, the rapid changes that the wave shape and spectral content undergo in a short time result in a high rate of information delivered to the perceptual system. In contrast, the flow of information saturates once the steady state is reached. For example, for identification of clarinet sounds it was found that scores do not improve after 0.5–0.6 s of presentation time [32.180]. In regard to detection and identification of onsets and transients as well as of changes in spectral envelope over time, there are temporal and dynamic thresholds as well as several integration constants that have been reported from experimental findings (for an overview of research, see [32.31]). Further, there are forward and backward masking effects (for a detailed discussion, see [32.181]). Some of the relevant integration constants concern the very limit of auditory temporal discrimination between two events and the threshold of temporal order between events. The limit of temporal discriminability has been given as 3–5 ms (as in experiments on gap detection), however, the absolute value depends somewhat on sound level and other conditions. Other time constants concern just noticeable differences in onset asynchronicity and the threshold of temporal order. Onset asynchronicity between instruments playing notes simultaneously typically is in the range of 0–20 ms; asynchronicity within this range (and approaching its upper limit) supports identification of instruments [32.182]. Onset, in this respect, means the sensory and perceptual *event* that of course is dependent on the vibration pattern of sounds yet is not identical with the SPL measured at a certain time. The perceptual onset for tones in succession will be detected when the amplitude of vibration and the SPL of the sound radiated from a source exceed a certain threshold, which has been given as 6–15 dB below the maximum level of a tone [32.183]; this threshold is variable with respect to the absolute SPL of the stimulus and is also variable due to masking effects if several tones/sounds are played in close succession or nearly simultaneously. Two tones played in succession become clearly discriminable as *auditory events* when their perceptual onsets are about 40 ms apart; the threshold for audible single reflections of sound from walls to be perceived as *echoes* is in the range of 40–50 ms so that several early reflections occurring within a shorter time span ($t < 30$ ms) typically *smear* into one sensation. Given that ≈ 50 ms mark the time difference between events to be perceived as an orderly sequence and that complex tones in general are identified in regard to their pitch, timbre and loudness in a time window of ≈ 100–200 ms, it follows that the number of notes in music that are clearly identifiable as a melodic and diastematic structure, plus conveying a certain timbre and loudness, is limited per time unit.

Onsets can be very short as in swinging and carillon bells where, after the energy is transferred from the clapper to the bell (contact time ≤ 1 ms [32.184]), a large number of modes is excited within a few ms since the wave speed in bronze is ≈ 4400 m/s for longitudinal and ≈ 2160 m/s for transversal waves. The sound radiated from the bell or from a similar idiophone struck with a hard mallet contains many strong harmonic and inharmonic partials as a complex mixture [32.85, 86, 185] from which the auditory system must derive pitch and timbre information [32.186]. The onset of such sounds (which fall into a time window of 20–50 ms) usually is perceived as *clangorous* due to considerable spectral inharmonicity as well as high-frequency content, and assignment of a single pitch to the onset segment often is not possible. If idiophones struck with a mallet are at one end of a scale measuring the acoustical transient portion of a sound, there are some instruments on the other where the transient is long and can last for several hundred milliseconds as, for instance, in a double-bass played softly with a bow [32.187] or in large organ flue and reed pipes [32.35]. Also, some folk instruments such as the Slovak duct flute *Fujara* (of ≈ 160 cm tube length [32.41]) are slow in the buildup of modes. However, the dynamics of the onset of a sound is much dependent on playing technique as well as the dynamics of the musical context in which instruments are used. In works of modern music, notation might prescribe dynamics from *fffff* to *ppppp* and might include verbal instructions that range from *tutta la forza* to *morendo*. Also, composers in the decades after 1950 demanded instrumentalists (and also singers) to produce sounds in often uncommon ways, which could bring about, for example, quite percussive sounds from wind and string instruments. The variability of onsets due to playing technique and context notwithstanding, subjects familiar with music and orchestral instruments make use of the onset of sounds as information bearing to the classification and identification of such instruments.

For listeners, the transient part of sounds serves to distinguish various instruments as well as to mark the onset of notes. Transients such as the *spitting* of organ pipes, the prerunner sound in harpsichord strings (when the plectrum touches the string before lifting and plucking it [32.188, 189]), the *raspy* attack of a bow set to

motion on a cello string (in particular the C- and the G-string), the plosive attack in a trombone or trumpet tone, are such *prefixes* (to use Schouten's term) to the steady state that make sounds characteristic of certain instruments or families of instruments. Some instruments also have a peculiar decay or have markers at the end of a tone; for instance, in most harpsichords one can hear the jack fall down when a key is released, an effect that is prominent when full chords are played. In an ensemble performance, the *prefixes* from individual instruments can help listeners to notice the onset of the tones they play (listen, for example, to Hindemith's *Kleine Kammermusik für fünf Bläser,* op. 24,2). If attack transients and the final decay to zero amplitude are removed from natural instrument tones, identification scores drop significantly even for musically trained subjects where the effect is greatest for the attack removed [32.190]. Conversely, in an MDS study [32.166], similarity ratings for onsets from natural instruments presented in isolation did not differ much from complete tones, indicating that the onset contains most of the relevant information [32.25]. Removing perceptual cues like the transient at the onset means an increase in confusability of stimuli. Apparently, the effect can be compensated, to some extent, by playing technique. For example, instruments typically played with vibrato (flute, violin) are less affected by such removal since the vibrato then may substitute the cue needed for identification [32.191].

Time (or Temporal) Envelope

Isolated sounds from various instruments can be distinguished with respect to their temporal envelope, which can be segmented into characteristic parts (as in the attack, decay, sustain, release (ADSR) model Sect. 30.2). For instruments excited by a single impulse (many idiophones, most membranophones) a rapid attack is followed by a fast decay. Large bells have a long sustain, however, when sound is radiated on a low level for tens of seconds. In aerophones, tones die away quickly once the air supply to the generator is stopped. In plucked and bowed chordophones, sustain can be quite long after excitation has stopped because parts of the resonator and the air enclosed in the box still vibrate (due to a storage of energy). For the tone G_2 played on the open string of a cello (staccato, forte), the sustain lasts for ≈ 3 s after the bow has been lifted from the string (Fig. 32.8; the sound was recorded ≈ 0.5 m away from the instrument in a *dry* studio).

In this tone, the rapid attack and the long sustain are perceptually salient envelope features. In general, there are certain shapes of envelopes which, combined with spectral patterns, probably make up *timbral prototypes* for listeners. Changing the temporal and/or the spectral envelope means part of the information does not match a certain type of instrument. For example, reversing a natural complex sound affects both cues since, even though the total energy contained in the sound is the same when played forward or backward, the order in which relevant features become audible is reversed, which makes identification of natural sounds reversed in time difficult. The effect of time reversal of sounds was used quite extensively in the 1960s, in recordings of *psychedelic* music, where usually one electric guitar track previously recorded and then played backwards provides lead guitar lines in an otherwise normal mix (listen, for example, to The Byrds: *Thoughts and Words*, recorded in December 1966). An important factor for sensation and perception of temporal envelopes is modulation. Regular amplitude modulation (AM) can result from narrowly spaced spectral components, which would give rise to a sensation of roughness

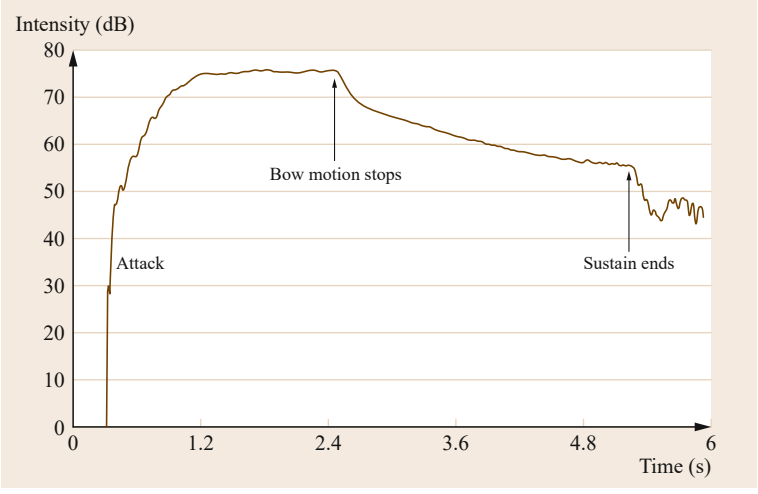


Fig. 32.8 Cello, open G-string (G_2), staccato, forte, steep attack, long sustain

(Sect. 31.6.2) in addition to the amplitude fluctuation. Quite regular AM with very little FM of f_1 and without introducing roughness can be produced also by modulating excitation parameters such as the blowing pressure in a flute [32.192].

Tonalness versus Noise

Tonalness usually is understood as a sensory attribute dependent on spectral harmonicity as well as its equivalent, periodicity of the time signal. Tonalness is one factor, besides roughness, sharpness (or stridency, Sect. 32.2.2: *Semantic Attributes of Timbre*), and loudness in a psychoacoustic model of *sensory euphony* [32.145, 146]. In regard to spectral harmonicity ≡ temporal periodicity (consequent to the Wiener–Khintchine theorem, Sect. 30.2), the steady state of a tone from an aerophone or chordophone played with medium force of excitation is basically tonal while the transient onset can contain noise in significant quantity. Some noise also results from the motion of the bow on a string or from the flow of air in wind instruments (there are playing techniques for flutes and saxophones emphasizing a *breathy* sound). In order to make synthesized sounds appear natural to listeners, their transient part needs to be shaped and noise must be added to sinusoidals in a harmonic or inharmonic complex [32.193, 194].

Spectral Envelope

Perhaps the most important component in timbre perception (and the defining criterion for *sound color*) is the spectral envelope. Additive or subtractive synthesis as was implemented in a host of electronic organs of the 1960s produces sounds which, though lacking characteristic onset and decay of natural instruments, at least are indicative of certain classes and types of instruments (those organs usually offered a selection of *flutes*, *reeds*, and *brass* sounds in more or less fair imitations of the originals). Spectral envelopes for these classes often were derived from filtering a complex (e.g., rectangular) waveshape so that concentration of energy in a band from 1−2 kHz resulted in a more nasal (*reed*) sound while emphasis on energy in low partials and in higher bands (2−5 kHz) should indicate a *brass family* sound. Flutes simply were imitated by low-pass filtering with the cutoff frequency set for a *mellow* sound. Since such synthetic steady-state sounds after a short while are experienced as *static* by listeners (for the lack of fresh information, see above), most electronic organs offered some AM (*Tremolo*) unit modulating amplifier output level while some had facilities to enrich the sound (for instance, by pairs of oscillators slightly detuned against each other so that organ stops with double sets of pipes such as *vox coelestis* or *unda maris* were imitated).

The concept according to which certain instruments or families of instruments distinguish themselves by spectral structure and the shape of the spectral envelope owes much to the source-filter model (Sect. 32.1.2). According to empirical findings [32.169, Chap. 6], timbre is essentially determined by the absolute frequency position of a spectral envelope, which suggests the perceptual attribute of timbre has a physical correlate in formant-like spectral energy distribution. Given that many instruments (aerophones including the singing voice, chordophones) are driven by pulse sequences fed into a resonator [32.195], the sound radiated from the instrument depends significantly on the geometry of the resonator. The size and shape of the resonator largely determines the spectral envelope (assuming the instrument is in its normal register, and played *mf*), which is sensed as a peculiar *sound color* [32.177]. This concept of approximately constant tone or sound color is behind organ pipe stops where different stops (or ranks) of flue and reed pipes distinguish themselves by their sound color, for pipes of the same pitch (determined basically by pipe length, measured in foot, e.g., 8′, 4′, 2′). To maintain a given sound color (e.g., diapason, salicional, flute, trumpet) over several octaves, organ builders follow certain mensuration rules [32.196, Chap. 3]. For example, for two flue pipes an octave apart, the ratio of the pipe lengths is 2 : 1 while the pipe diameters should have a ratio of 1.682 : 1 (and the cross-sectional areas should be in the ratio of 2.828 : 1). That is, approximate homogeneity of sound color from one pipe tone of a given rank to the next is achieved by means of scale factors [32.11, 196, 197]. Correct scaling ideally would produce almost identical spectral envelopes for all the tones within the gamut of several octaves; the envelope then simply is shifted along the frequency axis with rising f_1 of each tone. To illustrate the case, formant filter envelopes of three organ tones recorded from the same stop are shown in Fig. 32.9. The tones are C_2, C_3 and C_4 from a trumpet 8′ stop of the historical organ at Hollern. The envelopes are fairly similar given that the microphone distance relative to the three pipes was not identical and that some reflections of sound inside the organ case may have occurred (sound levels were normalized to −6 dB fs for analysis).

A scale factor that preserves the relations within a specific geometry can also be used for peals of (swinging or carillon) bells where the scaling of size and mass determines the pitches but should (ideally) not affect the sound color. Further, the same principle can be applied to *families* of instruments that cover different registers (soprano, alto, tenor, baritone, bass) but share the same basic sound color [32.195] in the steady-state portion of sound when playing conditions are kept almost constant. A violin, a viola, a cello and a double

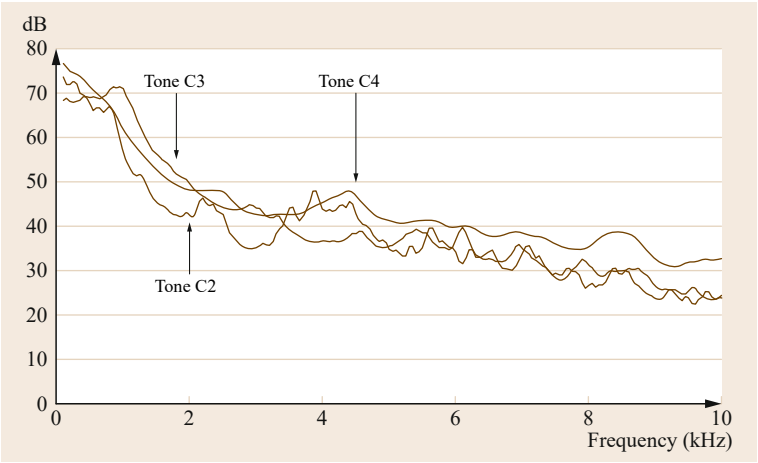


Fig. 32.9 Formant filter envelopes for three tones (C_2, C_3, C_4) of one organ stop (trumpet 8′)

bass are recognized as members of the *bowed strings family* notwithstanding differences in register and variation in spectral fine structure. If types of instruments can be distinguished (and possibly identified) by their respective spectral envelopes sensed as *sound color*, a hypothetical explanation is that the spectral energy distribution in the sound corresponds to a specific excitation pattern along the BM that is learned as representing a certain sound source, and is stored as a template or profile in long-term memory (LTM). However, one has to take the variation in sound color into account, which may occur in different registers of a single instrument, or even within one octave of its playing range. Also, factors such as the strength of excitation (usually from *ppp* to *fff*), dependence of spectral energy distribution and of the directivity pattern of radiation on sound level as well as room acoustics (absorption, reverberation) all influence the *timbre* one perceives of a given instrument [32.52, 198], [32.169, Chap. 6]. The problem of how homogeneous timbral qualities are relative to the tones of a musical scale within one octave, and more so if scales extend into another octave, has been studied empirically. Research on this topic involving MDS was done by *Marozeau* et al. [32.199] who concluded that pitch differences of tones within one octave had but little effect on timbre dissimilarity judgments. With respect to wind instrument sounds, data from musically naive subjects suggested timbre is perceived as identical for tones within one octave [32.200]. A replication of the experiment showed, however, that musicians can make reliable judgments beyond that range [32.201].

Change in Spectral Envelope and Pitch

A close inspection of many musical sounds reveals that both the period length of the complex waveshape and the frequencies and amplitudes of spectral components vary with time. For the steady state of complex harmonic sounds recorded from aerophones and chordophones played without vibrato, the variance of period length T (ms) or its inverse, the fundamental f_0, can be quite small, so that no pitch modulation is audible (in line with autocorrelation function (ACF) analysis). Pitch shifts, however, will be encountered in case tones are played with vibrato, which is customary nowadays for flute and violin performances, especially for the repertoire of the *romantic* era. Vibrato is also used extensively in belcanto singing (Fig. 31.27). Vibrato is performed, for instance on a violin or other bowed string instrument, rolling the cup of the finger up and down on a string whose vibrating length is thereby varied periodically. Vibrato thus produces FM; the modulation frequency usually applied by violinists is about 5−8 Hz and modulation of f_1 and higher partials can reach or even exceed ±35 cent. Vibrato can also give rise to AM of harmonic partials when their frequencies move in and out of the narrow resonance zones of the resonator [32.202]. Consequently, the spectral components in general show periodic AM along with the FM from the vibrato, however, individual components can be affected differently dependent on their position relative to the resonance zones. In sum, the pattern of FM plus AM resulting from vibrato may deviate somewhat from strict periodicity. To illustrate the case, a small segment from the recording of a professional violinist playing the note c#5 (over a chord provided from soft piano accompaniment) is shown in Fig. 32.10.

One can see that the violin partials undergo FM while they exhibit AM to different degrees; some of the tracks marking partial frequencies shift between strong and weaker amplitudes as indicated by black and gray color respectively. Players of the modern concert flute learn a special technique of breathing and chest muscle control [32.203] that enables them to produce FM vibrato and AM tremolo effects. Periodic variation of

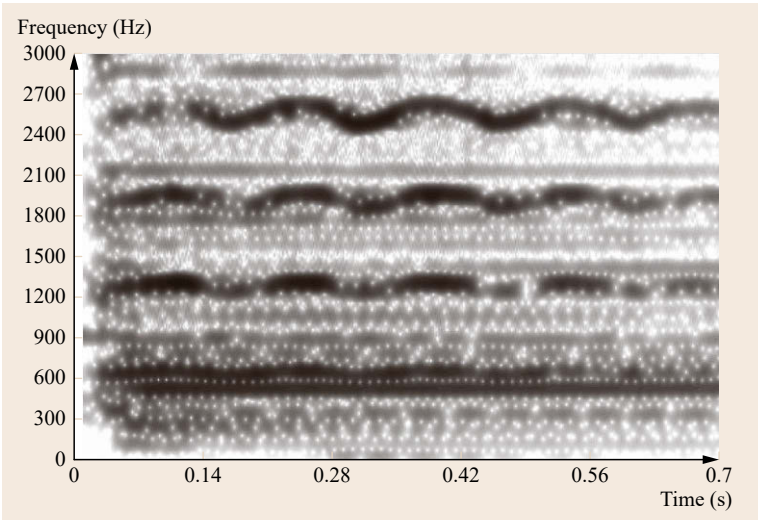


Fig. 32.10 Violin, note C#5 ($f_1 \approx$ 554 Hz), vibrato, FM and AM effects

blowing pressure results in a tremolo without significant FM [32.192]. Changes in blowing pressure means the force of excitation is varied over time, which results in different numbers of natural modes that are excited in the resonator. Also, amplitudes of partials contained in the sound radiated from the instrument vary as a function of blowing pressure. In general, higher levels at the input of the system produce more partials so that the spectrum broadens toward higher frequencies. Consequently, the spectral centroid also shifts upwards in frequency, which can be sensed as an increase in brightness. Further, subjects are sensitive to changes in timbre evoked by phase shifts between harmonic partials [32.162]. The effect is strongest for harmonic complexes where partials are either in sine or in cosine phase as opposed to partials alternating in sine and cosine phase; the strength of the timbral difference depends on f_1 of the complex and varies markedly between subjects. For $f_1 = 294.4$ Hz the maximal effect of phase on timbre for a sample of eight subjects was equal to changing the SPL by about 2 dB. A possible explanation is that the peak amplitude per period and the crest factor are higher for harmonic partials in cosine phase compared to partials in alternating phase. For harmonic complexes consisting of five partials with $f_1 = 200$ Hz and amplitudes decreasing by -6 dB with harmonic number, the difference is ≈ 1.7 dB.

Summing up this section, *timbre* was found to depend on several temporal and spectral properties of sounds, which often combine into complex spectrotemporal patterns that are analyzed into constituents in perception in order to distinguish individual instruments or other sound sources. Modeling sensory analysis as carried out in the auditory periphery, the concept of CBs usually is implemented by chains of bandpass filters suited to perform spectral analysis as a basis for perception of pitch (Sect. 31.5). Such an approach can be followed also for timbre and loudness though in particular the transient part of sounds requires high temporal resolution, which means a conventional Fourier-based analysis may fall short of the performance of the auditory system, which is superior in regard to the *uncertainty relation* $\Delta f \Delta t$ [32.204]. For the steady state of most sounds, spectral envelope and centroid have proved to be good descriptors of sensory attributes (see above and also *Barthet* et al. [32.205] for clarinet tones). Viewed in terms of CBs, patterns of spectral energy distribution code pitch and timbral information as well as loudness (*Pollard* and *Jansson* [32.206], Sect. 32.2.4 and Chap. 33). Applying time constants relevant for auditory perception, timbre can be approached as a sequence of *windows* or *frames* that contain spectral energy distributions. If there is not much change from one frame to the next, a more or less stable spectral profile evolves, which can represent a *sound color* that in turn may indicate a certain instrument or *family* of instruments [32.195]. However, identification of instruments or other sound sources is facilitated when temporal cues are offered along with spectral information. Experiments have shown that attack time is an important cue where either soft onsets or steep slopes (Fig. 32.8) help to categorize sounds. In addition, the overall shape of the envelope can indicate whether sounds are percussive rather than continuant. Further, modulation (AM, FM; regular, quasiperiodic, or irregular) can be used as a cue for timbre perception and categorization. Sounds undergoing modulation (AM and/or FM) might appear *raspy* or *blurred*; some sounds appear *shimmering* or *clangorous* or *ringing* due to spectral inharmonicity and modulation. The actual

effect on sensation depends on modulation parameters (depth, frequency) as well as on how many spectral components are modulated and if modulation is conjoined for these components or not.

32.2.4 Interrelation of *Pitch* and *Timbre*

In concepts of tone perception developed in psychophysics [32.74, Chaps. 9–11], [32.176, 207] the pure tone figures prominently since it allows the establishment of a close correspondence between physical and sensory magnitudes. In regard to pure tones, it is feasible to address pitch as depending on the frequency of vibration, and loudness proportional in some way to the vibration amplitude and intensity of radiated sound. Duration does not pose a problem assuming time is a linear process (equivalent to the constant motion of a mass point in 3-D-space [32.208]). This leaves sound color or timbre as the perceptual quality that relates to several physical parameters (Sect. 32.2.3). As has been argued above, sound color constitutes a quality complementary to pitch if one adopts the perspective of Helmholtz based on Fourier and Ohm, which allows the decomposition of a harmonic complex into the fundamental (f_1) carrying all or at least most of the pitch information, and the remainder of the spectrum conveying the sound color that may characterize a certain instrument or family of instruments. Such a view might hold for the steady state of a harmonic complex where f_1 is dominant and the amplitudes of the other partials roll off at a certain rate (dB/oct) in the spectrum so that the pitch is unambiguously related to the fundamental. However, there are musical sounds in particular from idiophones and also membranophones where this model fails to capture relevant structures. For example, in *Western* swinging bells and carillon bells there are some partials close to harmonic frequency ratios while other spectral components are clearly inharmonic ([32.185] and Sect. 30.2). Typically, complex bell sounds give rise to more than one spectral or virtual pitch [32.104, Chap. 11], [32.186], and these sounds often show marked AM due to interaction of narrowly spaced components. Spectral inharmonicity and spectral modulation are features also found in non-Western idiophones, in particular in Javanese and Balinese *gamelan* [32.85, 86, 209, 210]. The combination of pitch ambiguity, spectral inharmonicity and modulation, which is characteristic of sounds from bells, gong chimes and other metallophones, is not compatible with the additive (f_1 pitch + spectral sound color) concept sketched above for harmonic complexes. Rather, such sounds are sensed as spectrotemporal conglomerates that are not easily analyzable into constituents by ear even though musically experienced subjects can assign pitches to many sounds by singing or humming a tone (or several if they sense more than one pitch per sound). Also, subjects make comments on the *clangy* or *metallic* onset and the *shimmering* decay of such sounds, which thus can be described verbally in terms of timbral attributes. However, for many complex inharmonic sounds there is even less a demarcation between the perception of *pitch* and that of *timbre* than might exist for harmonic complexes.

The issue of whether pitch and timbre are two distinct or two interrelated qualities that subjects perceive when listening to sounds such as synthesized harmonic complexes or tones played on certain instruments has been discussed on the basis of empirical data [32.211–213], [32.104, Chaps. 10–12]. In regard to criteria elaborated by *Garner* [32.172] for *integral* and *separable* dimensions of perception, there seem to be indications for both points of view. From common experience, one could argue that musically trained subjects are capable of assigning two labels to musical tones presented in isolation, one denoting a pitch and the other denoting a timbre that is characteristic of a certain instrument or at least a type or a *family* of instruments. The two judgments implied in such a labeling task are *categorical* in that the stimuli have to be ordered into discrete categories. Such a task can be accomplished if the cues available to subjects intending to identify both the pitch (in terms of musical denominations, e.g., F_3, B_4) and the instrument type (e.g., tenor sax, trombone, cello, electric bass) are unambiguous and salient. However, in experiments on possible interactions of pitch, timbre, and loudness involving synthetic sounds, subjects seemed unable to attend to one dimension or attribute while disregarding the other (two experiments coupled timbre and loudness as well as timbre and pitch [32.211]). If the task is discrimination of small changes in either pitch (with respect to f_0) or timbre (defined by spectral centroid) of synthetic harmonic complexes, musically trained subjects showed smaller difference limens (DLs) for pitch than nonmusicians but were similar in their respective spectral centroid DLs [32.213]. In addition, performance differed significantly between congruent (changes in f_0 and spectral centroid were in the same direction) and noncongruent conditions (which, in natural sounds such as produced from wind and string instruments, is unlikely). While some experiments suggest pitch and timbre can be perceived as independent of each other [32.212], interference of pitch and timbre has been reported as well. A possible explanation for conflicting evidence may be sought in different experimental designs, stimuli, and tasks. If the stimuli are tones from familiar musical instruments (or synthesized tones close in timbre to natural sounds), in particular

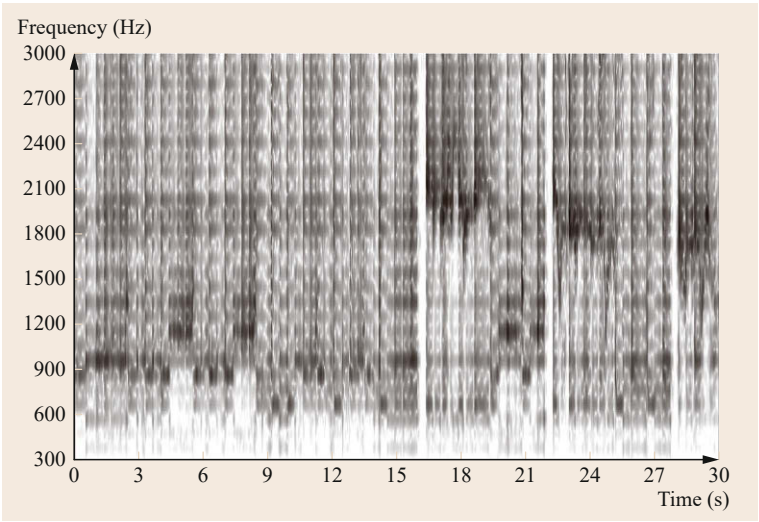


Fig. 32.11 Spectrogram 300 Hz – 3 kHz, Jew's harp melody, Opal Shuluu (Tuva)

musically trained subjects will have little difficulty in tracking both pitch changes and watching for changes in timbre (for example, shifts in centroid and brightness evoked through spectral filtering or by means of phase shifts between partials). Subjects thus can focus on one parameter at a time and quickly shift their attention between two parameters; such a strategy may be effective for essentially *independent* processing of pitch and timbre information. However, the structure of sound stimuli and of the sensory input used for processing has to be taken into account (Sects. 30.2, 31.1–31.6). For complex harmonic sounds such as radiated from aerophones and chordophones, BM filtering and auditory neural processing conveys pitch information both in place and in temporal code. Resolved partials and groups of unresolved partials all contribute to periodicity pitch while they also convey information concerning spectral energy distribution and spectral profile, which result in perception of sound color and the changes it may undergo in the course of presentation. Major changes in spectral structure and energy distribution can also have effects on perceived pitches. For instance, attenuation of the odd harmonics as well as phase shifts of partials in a complex can bring about a shift in perceived tone height by an octave or even several octaves while not affecting the *chroma* component of pitch ([32.214] and experiments reported in [32.215]).

The interdependence of pitch and timbre in harmonic complex tones is evident from music realized with instruments where a generator is put into the mouth that functions as resonator as well as with styles of vocal music where the mouth cavity serves as a bandpass filter. Such techniques can be observed in musical genres of various cultures where the Jew's harp (also: jaw's harp, trump) or the mouth bow are in use, or where styles of overtone singing are practiced. Figure 32.11 shows an excerpt of a melody played by Opal Shuluu (Tuva) on a Jew's harp [32.216, Track 31]. The spectrogram shows that each vertical sonority contains quite many components in the most relevant band (300 Hz–3 kHz), many of which are nearly harmonic and appear as multiples of a virtual pitch (autocorrelation, AC) at ≈ 96.1 Hz; the melody is filtered out by changes in the resonator (mouth cavity) so that certain components become more prominent in the spectral energy distribution per time frame. It is a typical mixture of melody against a drone, or, put in terms of Gestalt psychology [32.217, Chap. 7], of a figure against a ground.

With inharmonic sounds, separation of pitch and timbre is much more difficult since there is no strict periodicity relevant for f_0 pitch perception, and spectral structure may be also quite irregular. This hampers pitch perception for individual sounds from bells, gong chimes, or other metallophones [32.218]. Further, sequences of inharmonic sounds representing a scale or a melody may differ considerably in spectral energy distribution from one sound to the next as can be observed, for example, with bells in historic carillons. Spectrograms of sounds (cut to the initial 3 s) recorded from bells no. 1–3 of the Brugge carillon (Joris du Mery 1744, [32.186]) in Fig. 32.12 demonstrate that, notwithstanding essential components of a minor third bell that can be identified in all three segments, there is no homogeneous *sound color* since distribution of spectral energy varies markedly between these sounds. Further, the spectral component lowest in frequency (*arrows* in Fig. 32.12 marking the so-called *hum note*) in each of these three bell sounds is weak in intensity, which implies it will hardly elicit an individual spectral

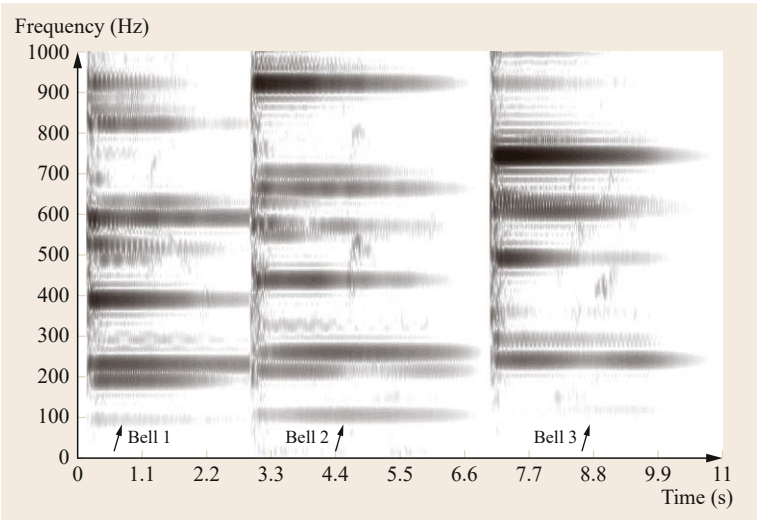


Fig. 32.12 Brugge carillon, sound segments from bells no. 1–3, spectral components 0–1 kHz, *fundamentals* (hum notes) marked by *arrows* are weak in these sounds

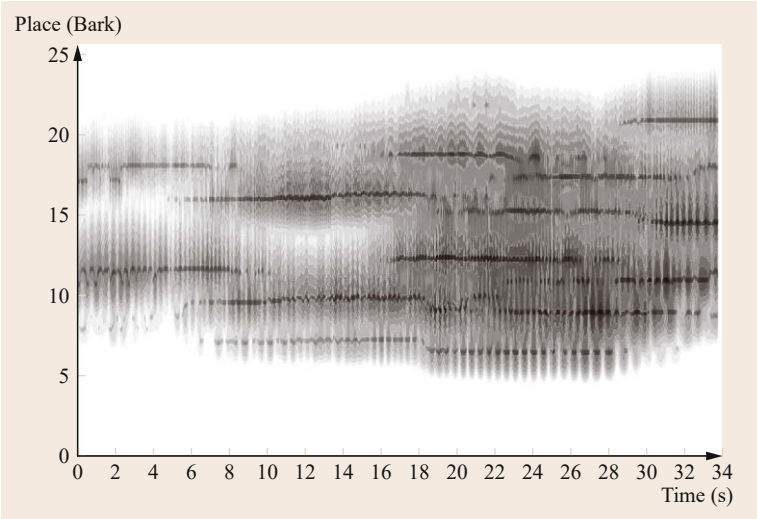


Fig. 32.13 John Chowning: excerpt from *Phoné*, left channel, cochleagram

pitch. Unlike the fundamental f_1 in a harmonic complex, which is reinforced by the sum of partials contributing to f_0 ($f_1 \equiv f_0$), virtual pitches in inharmonic complexes must not converge with the lowest spectral component. In bell sounds, the most prominent virtual pitch (the strike note) often is close to, but not identical with, the second spectral component. If this is fairly strong, it can be sensed as a spectral pitch besides the strike note. Pairs of adjacent spectral and virtual pitches cause ambiguity in perception.

The ambiguous nature of inharmonic sounds has been explored in compositions of computer music where textures of harmonic and inharmonic complexes are interwoven as in works of John Chowning who made use of the FM sound generation technique he had developed [32.59, 219]. In one such work, *Phoné* (1980/81), sequences of complex sounds, which at times resemble formant structures of the human singing voice, are distributed on two stereo channels (in the CD mix [32.220]). Though the listener can detect tonal elements in the complex sounds from both channels, spacing of spectral components is quite dense and energy distribution covers much of the audio bandwidth. A cochleagram (Fig. 32.13) shows an excerpt from *Phoné* (left channel only); one can see several strong components per time unit that are relevant in regard to spectral and/or virtual pitches interspersed with broad bands of energy. The perception resulting from both channels is a complex mixture of *timbral* elements with faint pitch structures in between. The vertical structure of these complex sonorities thus differs from traditional compositions where, typically, several *voices* can be

distinguished, which interact in polyphonic settings or combine into chord-like formations.

Electronic music based on analogue technology had already lifted the boundaries between *voices* and *timbre*. With digital technology, one can create complex sonorities of many kinds. Also, interpolation between various spectra chosen from natural sources can be effected (spectral interpolation is different from cross-fades between sounds as are available in most samplers, *Arfib* et al. [32.221]). The approach to sound generation based on digital FM opened new perspectives since, taking the basic concept of FM, that is, carrier and modulator frequencies, one can combine several modules that produce either carrier or modulator frequencies in a network that can generate arbitrary spectra, which again are modulated in frequency (or in phase) and in amplitude over time. Such sounds in fact can contain several inharmonic and/or harmonic complexes at any given time. Implementation of this concept became available in digital synthesizers suited to daily use in the studio, and even on stage, in 1983 (DX 7 and follow-up models like TX 802, TX 81Z). With digital FM sound generation one can set parameters so that octaves in a scale do have frequency ratios other than 2 : 1 (e.g., the *golden mean*, 1.618 : 1, which Chowning employed in *Stria*, 1976/77, [32.63, 219]). Further, FM technology enables composers to create scales, melodic sequences and chord-like sonorities that appear paradoxical in certain respects, and may be perceived like *auditory illusions* [32.222]. In compositions like *Stria* and *Phoné*, textures and mixtures of complex sounds may still evoke perceptions of *pitch* and *timbre*, however, the flow of sonic objects that are heard, moreover, moving in a 3-D space (the original version of *Stria* is quadraphonic), transcends traditional categories of *tone*, *pitch*, *voice* (in the sense of harmony and counterpoint), and *timbre*.

Composing with complex sounds rather than with musical tones is an approach realized in electronic and computer music as well as in such orchestral works where *clusters* of tones are equivalent to complex spectral structures. In regard to perception and musical syntax, an issue much debated is whether sequences of complex harmonic and/or inharmonic sounds may constitute a *scale* similar in function to a scale of pitches we perceive when hearing a sequence of harmonic complex tones. Some exploratory studies suggested a hierarchical organization of timbre similar to that of pitch would be feasible, at least in principle [32.164, 165, 223].

The issue whether sound sequences could be constructed in which shifts in timbre is the parameter equivalent to pitch shifts like in a musical scale, apparently was nourished from some sketchy ideas on the possibility of a *Klangfarbenmelodie* that *Schönberg* [32.224, p. 471] had added to his textbook of harmony. Schönberg seems to suggest that sounds might be varied in *sound color* so that identifiable sequences more or less analogous to pitch sequences would result. Schönberg did not go into detail except noting that, as he saw it, *Klangfarbe* was a more general concept comprising *Klanghöhe* as one dimension, referring to a sensation of relative height evoked by one complex sound when compared to another. His idea then was to vary *Klanghöhe* in such a way that sequences similar to a melody would be perceived. To be sure, *Klanghöhe* is not identical with *tone height* defined by linear frequency in two-componential models of pitch. A likely interpretation of Schönberg's short remarks is that *Klanghöhe* can be taken as equivalent to the spectral centroid. Consequently, operations on sounds that would shift the centroid up and down like on a pitch scale while maintaining the shape of the spectral envelope might be suited to create a *Klangfarbenmelodie*. Schönberg's own approach to this concept as manifest in his op. 16/III was that he had five tones in a complex sonority changing so as to produce noticeable changes in brightness over time [32.225]. Sensory brightness of a sound is largely dependent on the centroid resulting from spectral energy distribution.

If the spectral envelope would be identical for all the tones played by one particular instrument in different registers (see above), a near-constancy of sensational quality could be expected for listeners as the spectrum is virtually shifted up or down in frequency without changing the amplitude relations between partials. There have been considerations of how operations such as transposition and inversion (known from operations on melodic pitch sequences such as canons) could be applied to sounds with respect to the sensory and perceptual quality of *sound color* [32.177]. The problem, however, remains that shifts of the spectrum, while maintaining a more or less identical sound color, might not be perceived as such, yet rather as interacting with pitch structures since pitch, in particular for musically trained subjects, appears to be the more fundamental perceptual quality. Consider for example the sounds from 31 diapason pipes per octave as are available on the organ that has been built for the Huygens–Fokker tone system [32.226]. Playing the tones of this quite unusual scale, one after another (the difference in f_1 between adjacent pipes is 38.71 cent), musically trained listeners will perceive a sequence of tones distinctive in pitch and almost homogeneous in sound color. If one conceives of timbral sequences based on complex inharmonic FM sounds (see above) meant to constitute a *timbre scale*, it is likely that listeners with a musical background may still be inclined to infer pitch relations from such sequences though they may also perceive

a concept of *order* among sounds varying in certain timbral parameters.

Of course, in music and musical instruments there are interrelations between pitch and timbre in many respects. For example, mixture stops in pipe organs were introduced to expand the spectral width and to strengthen the brightness of organ sounds as well as to reinforce the pitch of tones that were played with fundamental stops (such as the diapason or Prinzipal). There is a combination of spectral and virtual pitch effects if the tuning of the keyboard matches that of pipes in mixture stops as close as possible. Also, interrelations of pitch and timbre have played a significant role in regard to musical composition and orchestration since composers knew from experience which instruments would *fuse* well in a homophonic texture and which instruments would be suited to support perception of individual voices in a polyphonic setting. Viewed from acoustics and psychoacoustics, there are factors such as formant-like concentration of spectral energy, partial spectral masking between instruments and sensation of roughness caused by spectral interaction in dissonant sonorities that need to be considered [32.181, 227]. Spectral fusion versus spectral roughness as well as emphasis of formants in singing styles are also relevant in vocal music, as for example folk music idioms can differ significantly in this respect [32.228].

32.2.5 Sound Segregation and Auditory Streaming

The performance of the auditory system in mammals is striking in several respects:

1. Sensory processing is fast and quite precise with respect to locating sound sources and extracting features from complex sounds.
2. The auditory system is capable of integrating related information into entities that become perceivable as *objects*.
3. The auditory system can distinguish between several concurrent sound sources and objects so that these can be identified and categorized accordingly.

This section will provide basics on hearing conditions in environments and will then survey some of the principles underlying formation of auditory objects, on the one hand, and their segregation when occurring simultaneously, on the other. Since in particular auditory stream segregation has been the subject of comprehensive monographs [32.217, 229], the following paragraphs will only cover some of the relevant points.

Listening to music in a concert hall or in front of an audio system (which may be stereophonic, quadraphonic, or ambiophonic) means a quasicontinuous stream of sound waves propagating through a medium (Sect. 30.3) reaches both ears of a subject who may try to identify sound sources and *objects* within this stream. In this respect, binaural hearing is the normal situation. Depending on the environment (which may be a concert hall, an open air concert, or one's living room), the ratio of sound energy emitted directly from the source and the sound reflected from hard surfaces may vary. In open spaces unbounded by reflecting surfaces, a free field condition prevails. If music is presented on stage in a concert hall, listeners typically have the orchestra or band in front so that most of the sound energy is transmitted directly, with a certain portion of lateral energy reflected from the side walls; in addition, energy might be reflected from a hard ceiling (for room acoustics and their effects on auditory perception, see [32.230–232]). In effect, in rooms bounded by hard surfaces, there is a mixture of direct sound and diffused sound between which a delay can be measured. The interaural cross-correlation [32.230, Chap. 3] between sound signals fed into both ears is a parameter suited to measure the degree of diffuseness. There are several more parameters (such as clarity, reverb time, coloration, distinctiveness or *definition*) that are of relevance for perception. For sound sensed binaurally in a free field (or in other spaces with negligible reverberation), mammals can use the interaural time difference (ITD) and the interaural level difference (ILD) as primary cues for spatial hearing and source localization [32.233, 234]. A model widely accepted is that the interaural time difference (ITD) is processed at the level of the inferior colliculus (IC), in pooled neurons [32.235]. Acuity is extremely high in that the just-noticeable difference (JND) for ITD seems to be close to 10 µs for a 500 Hz tone. In addition, the interaural level difference (ILD) serves as a cue for localization [32.236]. Though both ITD and ILD are restricted to certain conditions and ranges, taken together they can provide sufficient information to subjects for localizing sound sources. In regard to prerecorded music reproduced from audio systems, listeners are mostly confronted with a stereophonic setup where sources are *panned* on a left–right axis while 5.1 and other surround sound systems simulate a 360° panorama. Wave field synthesis [32.237] can even improve spatial representations of sources from prerecorded music that are perceived as if distributed in a natural 3-D environment.

Patterns of sound waves entering the auditory system binaurally can be extremely complex according to musical and physical parameters encoded. Consider, for example, performances of symphonic works rich in harmonic textures and instrumentation where also the dynamic range may vary considerably over time. Hence, listening to music requires fast processing

with sufficient resolution to allow for feature extraction and overall categorization of the input. Fast extraction of stimulus features necessary for sensory and motor responses in afferent-efferent feedback loops affords distributed processing along stations of the auditory pathway (AuP) (Sects. 31.2–31.4). As expected, there are indications that much of the processing relevant for pitch and other features is done subcortically, and in parallel up to the IC [32.238–240]. However, if processing is hierarchical and distributed, one would expect some neural network capable of integrating related information so that one perceives *objects* or even *complex wholes* and not just a bunch of features. This problem has been discussed quite extensively, in psychology and neuroscience, as one of *binding* [32.241, 242]. Though many studies on *binding* are concerned with visual perception as well as with language, formation of auditory objects also calls for a neural system suited to integrate spatial and temporal information gained from the various stages of analysis in the cochlea and along the ascending AuP. In the following, temporal and spectral criteria relevant for *fusion* as well as for *fission* of components in individual sounds and for fusion or segregation of concurrent sounds will be reviewed. Cues for identification of sound sources such as instruments and voices as well as for perceiving sonic objects embedded in a quasicontinuous stream of waves are temporal, spectral, and dynamic.

Fusion and *Fission* of Spectral Components in Individual Sounds

Evidently, we can hear a harmonic complex tone played on an aerophone or chordophone (e.g., oboe, trumpet, cello, sax) as a coherent whole notwithstanding that a number of low partials of such tones will be resolved in the BM filter bank, and also groups of higher partials will be segregated according to the CB auditory filter model [32.243–245] In a nonanalytic listening situation, individual harmonic partials and groups of partials falling into different CBs will be perceived as *fused* into one complex that, because harmonic partials join into a common period, gives rise to a distinct pitch and can convey a sensation of a certain timbre. The section of a sound that appears as *fused* into one object is the steady state while at the onset individual harmonics may be audible because, in particular in aerophones, some modes can reach a stable regime of vibration earlier than the bulk of modes making up the spectrum [32.204]. Of course, one can adopt an analytic stance and try to *hear out* some of the low partials in, for example, a harmonic complex comprising ten partials ($f_1 = 200$ Hz) with amplitudes rolling off at 3 dB/oct. Also, a nonharmonic component included in a harmonic spectrum most likely will be detected (if strong enough in level) as a separate component not fitting to the main body of partials constituting the perceptual object (a complex harmonic tone). Consider, for example, a harmonic spectrum where the third partial is detuned to a frequency ratio of 2.55 to $f_1 = 100$ Hz while the other components are in small integer frequency ratios. Using ten components with amplitudes decreasing at -3 dB/oct, the waveshape plotted in Fig. 32.14 results. The basic periodicity of 10 ms corresponding to f_1 as well as to f_0 resulting from partials 1, 2, 4–10 is still dominant; however, the inharmonic component obstructs a regular waveshape to repeat per period.

Separation in this case is effected by means of two concurrent pitch percepts, one based on the spectrum and periodicity (f_1 and f_0) of the harmonic complex, the other on the frequency and period of the inhar-

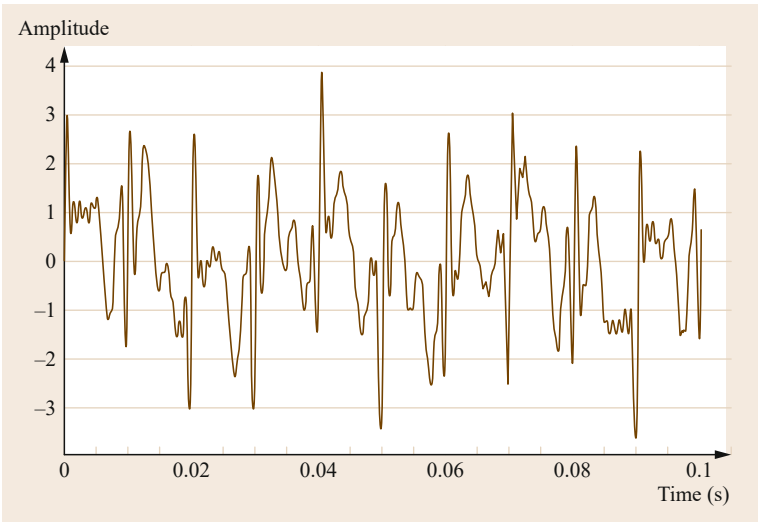


Fig. 32.14 Waveshape, harmonics 1, 2, 4–8 of $f_1 = 100$ Hz; partial no. 3 detuned to 255 Hz

monic component that *stands out* against the complex. Such a situation basically occurs with bell sounds where several partials may *fuse* quite well while in particular the minor third often is clearly detectable as a strong component (Fig. 32.12). Most sounds from carillon and swinging bells as well as such produced from gong chimes or by means of FM technique (see above) are ambiguous not because they would lack pitch and timbre information but because this information does not integrate easily into representing a single, coherent sound object. Rather, subjects in experiments tend to assign two or even more pitches to one complex inharmonic sound and in addition may find the timbre variable with time [32.85, 86, 209, 246–248].

Onset Synchrony versus Asynchrony of Concurrent Parts and Sounds in Music

Theoretical onset times and durations of notes in works of music can be calculated from notations in case a certain metronome value is prescribed. Absolute tempo (measured in beats per minute, BPM) in many recordings of pop music is evident from drum computer or sequencer tracks. In music performances not bound by notation and/or absolute tempo set by a machine, onset times of various instruments may be measured relative to some *time keeper* (for instance, a regular sequence of fast notes the drummer plays on the ride cymbal). Jazz and also rock musicians seem to have opinions of what may contribute to a good *groove* in a musical performance. One aspect often alluded to is that onsets of certain instruments should be slightly ahead or behind those of the time keeper. For instance, the bass may *drag* the tempo a tiny bit ahead of the time keeper while the snare drum is a tiny bit late, and then is perceived as somewhat *heavier* relative to the beat marked by a regular sequence of pulses (e.g., the attack of strokes on the ride cymbal). In such perceptions, two factors seem to be of relevance. One relates to temporal order and precision relative to an audible or imagined pulse, the other perhaps to effects of masking. If all onsets were synchronized so that sounds from different instruments in an ensemble would converge as much as possible, it would be difficult to distinguish their notes, in particular if instruments similar in sound color (see above) are playing notes in consonant chords. Of course, this is a situation wanted in certain musical contexts such as homophonic settings where maximum fusion of parts (as well as of partials) is desired. However, in polyphonic or multipart music, it is often necessary for listeners to follow individual parts (or *voices*) in order to apprehend musical structure as based on themes, motives, and voice-leading. For example, many works written for string or saxophone quartet require listeners capable of perceiving individual parts as well as the interplay of such parts in simultaneous chords or other vertical sonorities. Experimental data suggest that even musically trained subjects have difficulties in keeping track with multipart music if the number of voices exceeds three and when timbres for all voices or parts are relatively homogeneous [32.249]. Since timbre may not suffice to distinguish sources within *families* of instruments such as bowed strings or saxophones, and partial spectral masking can occur in particular if parts are relatively close in the pitches of their respective notes [32.181], temporal and dynamic factors come into play as a means to keep parts or voices apart. As has been reported by *Rasch* [32.182], subjects showed higher scores in correctly detecting the direction of intervals in quasisimultaneous concords formed of two complex tones when their onsets differed by 0–20 ms. *Rasch* [32.250] also tested the role of onset asynchrony in small ensembles where he found that those instruments that play the main melodic line tend to lead by about 30–50 ms. Onsets as in performed music seem to vary considerably relative to a grid one may calculate from notation, or may impose from a reference instrument. Onset asynchrony helps the listener to segregate voices in polyphonic music. In polyphonic keyboard works of the Baroque era (such as fugues written by D. Buxtehude, N. Bruhns, J.S. Bach), different voices (assigned to the right and left hand of the performer respectively) often do not begin aligned but with one voice starting on the beat and another kept apart by a quaver or semiquaver rest preceding the entry of that line.

Coordinated Modulation of Spectral Components as Marker of a Source

There are many reports on the effects modulation has on source segregation and identification. In particular, coordinated modulation of spectral components so that their frequencies vary in parallel has been stressed [32.251]. In experiments with synthesized vowels presented in combinations at different pitches, subjects judged the prominence of a modulated target vowel higher than unmodulated vowels [32.252]. In a musical situation such as when a solo violinist is backed by an orchestra in a violin concerto, there might neither be much difference in averaged (root mean square, rms) sound level between the solo violin and the orchestra nor in the timbre of the solo violin as compared to the string section of the orchestra. Parameters suited to effect acoustic and auditory segregation of the solo violin against the orchestra then can be strong onset attacks in nonlegato phrases and the use of substantial vibrato in legato phrases comprising long-held notes. In fact, this is observed in many performances. Detection of modulation has been found an efficient cue for computerized scene analysis [32.253].

Segregation of Harmonic Compounds as in Single Consonant Chords

As has been explained in Sect. 31.6.4, there are certain conditions for perceiving harmonic complex tones and combinations of such tones in terms of consonance, fusion, and *Verschmelzung*, which means sounds such as Stumpf's *ideal concord* (Fig. 31.15) are perceived as highly coherent while being also apprehended as a configuration of tones related by certain intervals. In order to apprehend structural relations between several pure or complex tones, a listener should be able to segregate a chord or other harmonic compound into its constituents. This requires that a sound is present for a certain time, and that the listener has some experience in analytic listening. The task is to find how many pure or complex tones are contained in a chord or other sonority, and to identify the relations between the components. Consider for example the final chord in a work of organ music like the *Praeambulum primi toni a 5 in d* by Matthias Weckmann where five voices join into a long-held chord comprising the notes D_2, D_3, A_3, $F\#_4$, A_4, D_5 (the note D appears tripled to emphasize the key of the piece). Since works like this *Praeambulum* are usually played with several stops at $16'$, $8'$ and probably also at $4'$ and $2'$, spectral energy relevant for pitch and timbre perception as contained in many partials should cover a frequency range from D_1 to at least three octaves above D_5, that is, ≈ 4.7 kHz. In the recording used here for analysis (Wilde–Schnitger organ of 1683, Lüdingworth near Cuxhaven), one of the stops is a $16'$ dulcian reed pipe that produces very many harmonic partials per tone. Figure 32.15 shows a spectrogram 0–2 kHz and the f_0 of the chord derived from AC and subharmonic summation (SHS) analyses as well as the output of a Bark filter analysis with the center frequencies spaced closer (ratio 0.85) than in the usual Bark scale to emulate CB bandwidths as observed in auditory peripheral filtering. These data-driven analyses demonstrate that segregation of constituents according to features (pitch-relevant partials, other harmonics, onsets, temporal fluctuations of spectral components in frequency and amplitude) is possible, at least to a certain extent.

Hearing the chord live in the church, one may try to use spatial information since the pipes sounding the chord are distributed in the organ (with the pedal stops housed in two towers flanking the main organ case). In regard to pitches, which are often the strongest cue for auditory analysis, one can hear the *fundamental* at $D_1 \sim 38.8$ Hz (the organ is tuned to $A_4 \sim 466.3$ Hz), which is the f_1 spectral component of two of the $16'$ pipes employed as well as a periodicity pitch (f_0) that is confirmed by both AC and SHS analyses (Fig. 32.15). Of the remaining tones, several can be identified by their pitch intervals relative to D_1, especially A_3, $F\#_4$, and D_5. Neglecting small fluctuations in frequency and amplitude (Fig. 32.15), the D-major chord to one's ear offers a high degree of *Verschmelzung* notwithstanding small tuning deficiencies (the organ is in meantone tuning, implying the $F\#_4$ is a just major third but the A_3 and A_4 are flat by 5.5 cent respectively). Auditory analysis in this case is not too difficult because the chord lasts for almost 5.5 s.

Segregation of Auditory Streams

In 1649 and 1654, *Jacob van Eyck*, a famous Dutch carillonist and flutist, published two volumes of a collection of some 150 tunes, many of which were part of the popular repertoire of his time [32.254]. These tunes, which can be performed with a single soprano

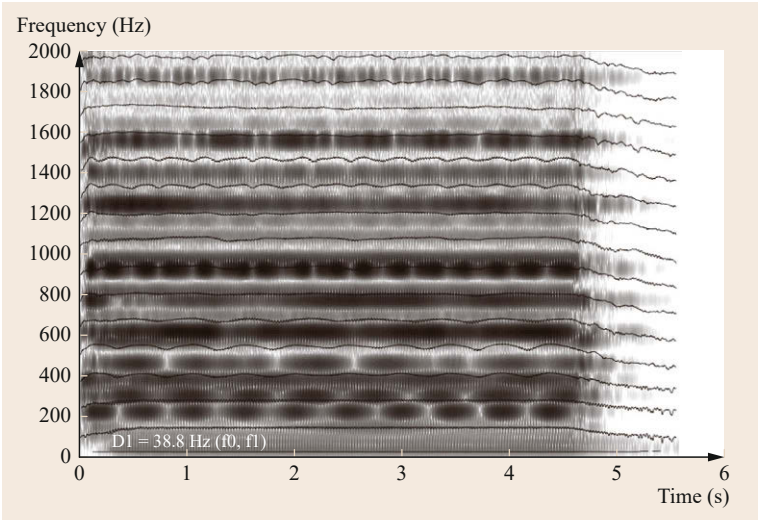


Fig. 32.15 M. Weckmann, *Praeambulum*, final D-chord, spectrogram 0–2 kHz, AC and SHS pitch analysis, Bark filter analysis 0–2 kHz (15 bands), extraction of f_0 (identical with f_1), detection of strong components and of (frequency, amplitude) modulation effects

recorder, are elaborated in variations, which include embellishments such as figurative elements as well as short notes inserted into the somewhat longer notes of the tune. In effect, if played at a lively or even fast speed by a skilled flutist (as van Eyck was), the impression on a listener is that the music, though not truly polyphonic (which would rather call for two or more independent voices), is not monophonic either. The type of setting found in van Eyck's anthology has been labeled pseudopolyphony; it was a technique employed also by J.S. Bach in his works for violin solo and cello solo (BWV 1001-1006; 1007-1012) where, however, simultaneous intervals and even full chords are included as these are playable on a bowed string instrument that offers four strings. In Fig. 32.16, a small section from the performance of one of the tunes adapted by van Eyck (*Wat zalmen op den Avond doen? Van Eyck* [32.254]; Marion Verbruggen, recorder) is shown, which demonstrates the style of a pseudopolyphony where two voices played on a single recorder seem to interlock in time.

As Fig. 32.16 demonstrates, most of the main notes making up the tune are in a higher register, and are played with more force than the very short notes that fall between them. Thus, there are two frequency regions divided at ≈ 950 Hz; the very first note/tone (~ 841 Hz, A_5^b) can be related to both the upper and the lower stream, which are divided by about an octave (tone no. 2 is at 1109.3 Hz, D_6^b, tone no. 3 is at 552.4 Hz, D_5^b, etc.).

The perception of pseudopolyphonic structures such as found in works of van Eyck and Bach depends on several parameters. One is the average interval between the upper and the lower sequence of tones, another is the tempo (measured in MM or BPM) of the performance, which determines the number of events per time unit (event density) as well as the relative duration of tones (e.g., quavers, semiquavers) to be played in a phrase. Musically trained listeners will be able to follow both the tune melody and the up-and-down motion of pitches installed between notes/tones of the upper and the lower register even if their duration can be quite short ($t \leq 200$ ms). If the tempo increases further (which can be done by digital time compression without affecting pitch) to 150% of the original version, it is more difficult to follow the pattern of up-and-down intervals. Doubling the tempo, and hence the presentation rate, hampers interval perception and leads to a percept where the tune is still present but the tones in between no longer join into kind of a *counter melody* (as is indicated by tones in the lower region of Fig. 32.16). Also the rhythmic pattern seems to have changed to a *galloping* type as has been described, for certain tone sequences, by *van Noorden* [32.255] and *Bregman* [32.229]. Of course, such a tempo would not suit any musical performance of this style of music. However, there are idioms such as the Kiganda Amadinda xylophone music of former Buganda [32.256, 257] where basically two melodic sequences are played simultaneously, in isochronous pulses, with no metric accent, at a very high speed (eighthnote $\sim$ 520–600 MM). Each sequence is of a certain length (for instance, 36 notes), and is repeated over and over. Unless one is familiar with Luganda language and the concept behind the music, it is not possible to apprehend individual sequences (which are mostly derived from songs that can be narrative in character). Rather, in particular *Western* listeners tend to perceive a number of patterns that have been explained as *inherent* in the fast tone sequences one actually hears [32.256, 258]. In this respect, such

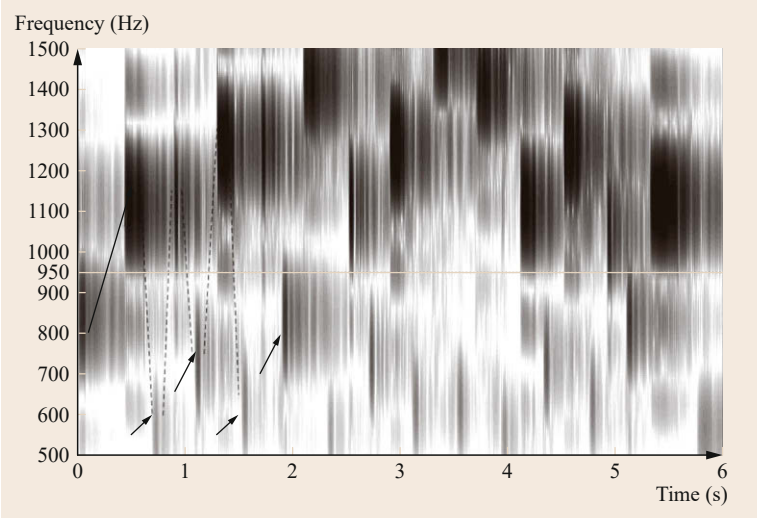


Fig. 32.16 Pseudopolyphony (after [32.254]); performance on a single recorder. The *beige line* at 950 Hz segregates two melodic streams. The tones in the lower stream (of which the first four are marked with *arrows*) are very short. The *solid line* connects the first two tones as were played and the *dashed lines* connect the tones 2, 3, 4, 5, 6, 7 in the up-and-down sequence. The decay in the tail end of the tones results from reverberation in the recording as well as from the filter response in the analysis

patterns have an objective basis. However, the Gestalt-like formations listeners perceive or imagine result from perceptual and cognitive structuring that leads to both segregation of tones within sequences into high and low pitch *streams* and to recombinations of elements into melorhythmic patterns. Since there is no objective rhythm implemented in isochronous pulse sequences produced without metric accents, the rhythmic grouping listeners realize seems to be derived from melodic patterns. The patterns listeners believe to perceive can be viewed as emerging from the very fast sequences, in an effort to group elements into coherent, musically probable formations.

The facts concerning pseudopolyphony and perception of inherent or emergent patterns briefly summarized in this section have been tested, on a much more fundamental level in a lab situation, in experiments on auditory stream segregation. An early experiment [32.259] found that, in the human hearing range up to ≈ 4 kHz, two pure tones A and B presented binaurally, each lasting 80 ms (plus 20 ms rise and decay time), appeared as a quasicontinuous up-and-down movement when their frequency difference df/f was within $\approx 15\%$ but turned into an interrupted two-tone pattern (A:B:A:B...) when the difference was larger. Seeing that the repetition rate for a pair A:B in this experiment is ≈ 5 Hz; the *trill threshold* – as the critical frequency difference was called – for 500 Hz would be close to 75 Hz, which equals 242 cent; as one might expect, this is about the width of a CB. *Bregman* and *Campbell* [32.260] reported two experiments one of which had six sine tones (2.5, 2, 1.6 kHz, 550, 430, 350 Hz) each lasting for 100 ms as stimuli. These tones were arranged in two sequences of high and low. With short duration of each tone, and high repetition rate of the sequences, subjects grouped the tones into two streams representing high and low tones. *Van Noorden* [32.255] conducted several experiments on sequential coherence of tones (as in melodies) and found that stream segregation depends on:

- The interval in pitch between two tones.
- The repetition rate. With increasing repetition rate, the interval width necessary for segregating tones at two different pitch levels into streams decreases.
- The intensity level differences between tones presented with alternating loudness, which can have an effect on perceptual grouping (termed *roll effect* by *van Noorden* [32.255]).

In addition, differences in timbre have been reported as a factor relevant in forming streams [32.261]. Finally, manipulating the phases of complexes of unresolved harmonics can lead to sequences of sounds that do not differ in power spectrum yet appear different in perception; such differences in perceptual quality may suffice for also inducing segregation [32.262].

Auditory stream segregation apparently is based on data-driven bottom-up analysis that starts at the auditory periphery. However, there are also top-down processes since segregation draws on learned schemata and on musical experience in general. It is of interest to note that quite many observations on auditory stream segregation can be viewed in regard to principles of Gestalt perception such as temporal and/or spatial *nearness* or proximity of elements, their *good continuation* as in a melody, the *common fate* partials of a harmonic complex share in FM, the *conciseness* of a certain rhythmic or tonal pattern, or its *closure*, etc. (a list of more than 100 principles of Gestalt perception was assembled by *Helson* [32.263]). A detailed discussion of experimental findings and an interpretation of many observations in terms of Gestalt perception and cognitive psychology will be found in [32.217] and [32.229]. Parallel to behavioral experiments, modeling of auditory stream segregation and development of algorithms for automated DSP-based analysis have been pursued. Though there are different approaches, many are based on peripheral auditory filtering of acoustic input and on emulating stages of neural processing [32.253, 264–267]. Separation of sources and assignment of sonic objects to *streams* or *voices* can be achieved by means of DSP algorithms operating on digitized sound files if sufficient information can be gathered from the cues named above (different onset times, different pitches and spectral patterns of concurrent sounds, identification of harmonic partials undergoing *common fate* FM as a marker of a common source, etc.). Another aspect that relates to auditory stream segregation and to auditory *scene analysis* in general is transcription of polyphonic music into conventional staff or other graphic notation [32.268–270].

Summing up this chapter on *timbre*, the term can be used to denote the spectral plus temporal features of a sound sensed by subjects as a tone quality that, however, may include dynamic changes. For example, the tone quality of a clangy metal gong sound resides largely in the modulation while the tone quality of a violin may be more dependent on spectral formant structure (and, hence, on *tone color*). In recent decades, it has been customary to address the phenomenal appearance of temporal and spectral sound features as they interrelate in sonic objects simply under the umbrella of *sound*. This is a term that has gained importance in particular in the production and perception of pop and rock music. *Sound*, in this respect, comprises various aspects having to do with the localization of sound sources in a stereo, quadraphonic or surround-sound mix as well as with the *depth* of space conveyed by the ratio of

the direct signal and (natural or artificial) reverberation. Original sound sources, in addition, can be modified by means of filters and compressors, and can be modulated in many ways. Characteristic of *sound* in pop and rock music genres is the use of special effects such as phasing, flanging, chorus, modulation of filter and spectral envelope parameters, cross-modulation between several sources, etc. applied to single or to groups of instruments (see articles by Dutilleux and Zölzer and by Evangelista in *Zölzer* [32.66], also [32.271]). Further, *sound* in pop, rock and also jazz music involves playing techniques (as is evident from rock guitar playing where one uses string bending, finger vibrato, so-called claw hammering, etc.) and also tuning of instruments. Note, for instance, the electric guitar on Pops Staples' *World in Motion* (Virgin, 1992) tuned low to C_2 instead of E_2. Moreover, the guitar apparently was played with an amp using a tremolo effect (often wrongly labeled *vibrato* in amp literature) and a 15″ loudspeaker to reinforce low frequency response and to make the sound appear *big* in volume. Further, some reverb has been put on the guitar (either in the amp or, more likely, in the mix). The *sound* resulting thus is a combination of the notes played plus the tuning of the guitar and the frequency response and other specifications of the technical devices used in the performance and recording of the song in question. *Sound*, in this respect, is a highly dynamical, time-variant construct that has objective physical and musical foundations. The actual interplay of temporal, spectral, spatial and dynamic features may undergo many changes in the course of a relatively short piece of music, which listeners may attend to; consequently, they will perceive *sound* as a dynamic process. However, one can also abstract the more recurrent and (within limits) more or less invariant features that are regarded *typical* of a certain sound (e.g., the types of distortion and of feedback in guitar sounds used in genres of hard rock productions).

References

32.1 D. Muzzulini: *Genealogie der Klangfarbe* (P. Lang, Bern 2006)

32.2 E. Dolan: *The Orchestral Revolution. Haydn and the Technologies of Timbre* (Cambridge Univ. Press, Cambridge 2013)

32.3 F.A. Gevaert: *Nouveau Traité d'instrumentation* (Lemoine, Bruxelles, Paris 1885)

32.4 H. Berlioz: *Traité d'instrumentation et d'orchestration* (Lemoine, Bruxelles, Paris 1904)

32.5 V. Mahillon: *Catalogue descriptif et analytique du Musée instrumental du Conservatoire royal de Bruxelles*, Annuaire du Conservatoire (Annoot-Braeckman, Gand 1878)

32.6 V. Mahillon: *Catalogue descriptif et analytique du Musée instrumental du Conservatoire royal de Bruxelles*, Annuaire du Conservatoire (Annoot-Braeckman, Gand 1879)

32.7 V. Mahillon: *Catalogue descriptif et analytique du Musée instrumental du Conservatoire royal de Bruxelles*, Annuaire du Conservatoire (Annoot-Braeckman, Gand 1880)

32.8 V. Mahillon: *Catalogue descriptif et analytique du Musée instrumental du Conservatoire royal de Bruxelles*, Annuaire du Conservatoire (Annoot-Braeckman, Gand 1881)

32.9 E. von Hornbostel, C. Sachs: Systematik der Musikinstrumente, Z. Ethnol. **46**, 553–590 (1914)

32.10 P. Morse, K.U. Ingard: *Theoretical Acoustics* (Princeton Univ. Press, Princeton 1986)

32.11 N. Fletcher, T. Rossing: *The Physics of Musical Instruments*, 2nd edn. (AIP/Springer, New York 1998)

32.12 C. Forsyth: *Orchestration*, 2nd edn. (Macmillan, London 1948)

32.13 B. Lewis (Ed.): *Bioacoustics. A Comparative Approach* (Academic, London 1983)

32.14 G. Manley, A.N. Popper, R. Fay (Eds.): *Evolution of the Vertebrate Auditory System* (Springer, New York 2004)

32.15 R.G. Busnel (Ed.): *Acoustic Behavior of Animals* (Elsevier, Amsterdam 1963)

32.16 G. Tembrock: *Biokommunikation. Informationsübertragung im biologischen Bereich*, Vol. 2 (Akademie-Verlag, Berlin 1971)

32.17 G. Witzany (Ed.): *Biocommunication of Animals* (Springer, Dordrecht 2014)

32.18 M. Konishi: Birdsong: From behaviour to neuron, Annual Rev. Neurosci. **8**, 125–170 (1985)

32.19 M. Naguib, K. Riebel: Singing in space and time: The biology of birdsong. In: *Biocommunication of Animals*, ed. by G. Witzany (Springer, Dordrecht 2014) pp. 233–247

32.20 L. Sayigh: Cetacean acoustic communication. In: *Biocommunication of Animals*, ed. by G. Witzany (Springer, Dordrecht 2014) pp. 275–297

32.21 J. Sundberg: *The Science of the Singing Voice* (Northern Illinois Univ. Press, DeKalb 1988)

32.22 N. Fletcher: Bird Song – A quantitative acoustic model, J. Theor. Biol. **135**, 455–481 (1988)

32.23 N. Fletcher: *Acoustic Systems in Biology* (Oxford Univ. Press, New York 1992)

32.24 G. Fant: *Acoustic Theory of Speech Production*, 2nd edn. (Mouton, The Hague 1970)

32.25 R. Bader: *Nonlinearities and Synchronization in Musical Acoustics and Music Psychology* (Springer, Berlin 2013)

32.26 A. Benade: On woodwind instrument bores, J. Acoust. Soc. Am. **31**, 137–146 (1959)

32.27 J. Roederer: *The Physics and Psychophysics of Music: An Introduction*, 3rd edn. (Springer, New York 1995)

32.28 J. Beauchamp: Analysis and synthesis of musical instrument sounds. In: *Analysis, Synthesis, and Perception of Musical Sounds*, ed. by J. Beauchamp (Springer, New York 2007) pp. 1–89

32.29 J. Fricke: Formantbildende Impulsfolgen bei Blasinstrumenten. In: *Fortschr. Akust. 4. Jahrestag. Akust. (DAGA), Braunschweig* (1975) pp. 407–411

32.30 W. Voigt: *Untersuchungen zur Formantbildung in Klängen von Fagott und Dulzianen* (Bosse, Regensburg 1975)

32.31 C. Reuter: *Der Einschwingvorgang nichtperkussiver Musikinstrumente* (P. Lang, Frankfurt am Main 1995)

32.32 E. Meyer, D. Guicking: *Schwingungslehre* (Vieweg, Braunschweig 1974)

32.33 S.L. Marple: *Digital Spectral Analysis. With applications* (Prentice-Hall, Englewood Cliffs 1987)

32.34 A. Schneider: Change and continuity in sound analysis: A review of concepts in regard to musical acoustics, music perception, and transcription. In: *Sound – Perception – Performance*, Current Research in Systematic Musicology, ed. by R. Bader (Springer, Cham 2013) pp. 71–111

32.35 A. Beurmann, A. Schneider, E. Lauer: Klanguntersuchungen an der Arp-Schnitger-Orgel zu St. Jacobi, Hamburg, Syst. Musikwiss. – Syst. Musicol. **6**, 151–187 (1998)

32.36 A. Beurmann, A. Schneider: Acoustics of the harpsichord: A case study. In: *Systematic and Comparative Musicology: Concepts, Methods, Findings*, ed. by A. Schneider (P. Lang, Frankfurt am Main 2008) pp. 241–263

32.37 D. Arfib, F. Keiler, U. Zölzer: Source-filter processing. In: *DAFX. Digital Audio Effects*, ed. by U. Zölzer (Wiley, Chichester 1996) pp. 299–372

32.38 X. Rodet, D. Schwarz: Spectral envelopes and additive + residual analysis/synthesis. In: *Analysis, Synthesis, and Perception of Musical Sounds*, ed. by J. Beauchamp (Springer, New York 2007) pp. 175–227

32.39 S. Tempelaars: *Signal processing, Speech and Music* (Swets Zeitlinger, Lisse 1996)

32.40 W. Hartmann: *Signals, Sound, and Sensation* (AIP/Springer, New York 1998)

32.41 O. Elschek: *Fujara. The Slovak Queen of European Flutes* (Music Centre, Bratislava 2006)

32.42 A. Beurmann, A. Schneider: Some Observations from a Stein-Conrad Hammerflügel from 1793. In: *Systematic Musicology: Empirical and Theoretical Studies*, ed. by A. Schneider (P. Lang, Frankfurt/M. 2011) pp. 175–184

32.43 P. Holmes: The Scandinavian bronze lurs. In: *The Bronze Lurs: 2nd Conference on the ICTM Study Group on Music Archaeology*, Vol. II, ed. by C. Lund (R. Swedish Acad. of Music, Stockholm 1986) pp. 51–125

32.44 L. von Falkenhausen: *Suspended Music. Chimebells in the culture of bronze age China* (Univ. of Cal. Press, Berkeley 1993)

32.45 M. Liang: *Music of the Billion. An Introduction to Chinese Musical Culture* (Heinrichshofen, New York 1985)

32.46 A. Barker: *Greek Musical Writings, Vol. 2: Harmonic and Acoustic Theory* (Cambridge Univ. Press, Cambridge 1989)

32.47 G. Wille: *Musica Romana. Die Musik im Leben der Römer* (Schippers, Amsterdam 1967)

32.48 M. Bröcker: *Die Drehleier. Ihr Bau und ihre Geschichte*, 2nd edn. (Verlag für Systematische Musikwissenschaft, Bonn 1977), 2 Vols.

32.49 H. Klotz: *Über die Orgelkunst der Gotik, der Renaissance und des Barock*, 2nd edn. (Bärenreiter, Kassel 1975)

32.50 M. Praetorius: Syntagma musicorum II: De Organographia (Holwein, Wolffenbüttel 1619)

32.51 C. Reuter: *Klangfarbe und Instrumentation* (P. Lang, Frankfurt am Main 2002)

32.52 J. Meyer: *Akustik und musikalische Aufführungspraxis*, 5th edn. (Bochinsky, Frankfurt am Main 2004)

32.53 W. Meyer-Eppler: *Elektrische Klangerzeugung. Elektronische Musik und synthetische Sprache* (Dümmler, Bonn 1949)

32.54 D. Ernst: *The Evolution of Electronic Music* (Schirmer, New York 1977)

32.55 Th Wells: *The Technique of Electronic Music*, 2nd edn. (Schirmer, New York 1981)

32.56 T. Holmes: *Electronic and experimental Music*, 4th edn. (Routledge, New York 2012)

32.57 A. Strange: *Electronic Music. Systems, Techniques, and Controls*, 2nd edn. (Brown, Dubuque 1983)

32.58 J. Watkinson: *The Art of Digital Audio* (Focal, London, Boston 1989)

32.59 J. Chowning: The synthesis of complex audio spectra by means of frequency modulation, J. Audio Eng. Soc. **21**, 526–534 (1973)

32.60 J. Chowning: The synthesis of complex audio spectra by means of frequency modulation, Comput. Music J. **1**, 46–54 (1977)

32.61 C. Roads, J. Strawn (Eds.): *Foundations of Computer Music* (MIT Press, Cambridge 1985) pp. 6–29

32.62 J. Chowning: Stria (1977), CD (Wergo/Schott, Mainz 1988)

32.63 B. Bossis: *Stria* de John Chowning ou *l'oxymoron* musical: du nombre d'or comme poétique. In: *John Chowning, Portraits polychromes*, ed. by M. de Maule (Éd. TUM, Paris 2005) pp. 87–113

32.64 P. Schaeffer: *La Musique Concrète*, 2nd edn. (Presses Univ. de France, Paris 1973)

32.65 P. Schaeffer: *Traité des objets musicaux* (Seuil, Paris 1966)

32.66 U. Zölzer (Ed.): *DAFX – Digital Audio Effects* (Wiley, Chichester 2002)

32.67 The Ronettes: Be My Baby (Philles Records, Los Angeles 1963)

32.68 H.F. Cohen: *Quantifying Music. The Science of Music at the First Stage of the Scientific Revolution* (Reidel, Dordrecht 1984) pp. 1580–1650

32.69 S. Dostrovsky, R. Cannon: Entstehung der musikalischen Akustik (1600–1750). In: *Hören, Messen und Rechnen in der frühen Neuzeit*, Geschichte der Musiktheorie, Vol. 6, ed. by F. Zaminer (Wissenschaftliche Buchgesellschaft,

Darmstadt 1987) pp. 7–79
32.70 J. Sauveur: *Collected Writings on Musical Acoustics (Paris 1700–1713)* (Diapason, Utrecht 1984), ed. by R. Rasch
32.71 J.P. Rameau: *Démonstration du principe de l'harmonie* (Pissot/Durand, Paris 1750)
32.72 F. Chladni: *Die Akustik*, 2nd edn. (Breitkopf Haertel, Leipzig 1830)
32.73 F. Opelt: *Allgemeine Theorie der Musik auf den Rhythmus der Klangwellenpulse gegründet* (Barth, Leipzig 1852)
32.74 E. Boring: *Sensation and Perception in the History of experimental Psychology* (Appleton-Century-Crofts, New York 1942)
32.75 R. Beyer: *Sound of our Time. Two hundred Years of Acoustics* (Springer/AIP, New York 1999)
32.76 C. Seashore: *The Present Status of Research in the Psychology of Music at the University of Iowa* (Univ. of Iowa Press, Iowa City 1928)
32.77 M. Metfessel: *Phonophotography in Folk Music. American Negro Songs in new notation* (Univ. of North Carolina Press, Chapel Hill 1928)
32.78 E. Meyer, G. Buchmann: Die Klangspektren der Musikinstrumente. In: *Sitzungsber. Preuss. Akad. Wiss., Math.-Phys. Kl*, Vol. XXXII (1931) pp. 735–778
32.79 H. Backhaus: Über die Bedeutung der Ausgleichsvorgänge in der Musik, Z. techn. Phys. **13**, 31–46 (1932)
32.80 F. Trendelenburg, E. Thienhaus, E. Franz: Klangeinsätze an der Orgel, Akust. Z. **1**, 59–76 (1936)
32.81 W. Graf: *Vergleichende Musikwissenschaft* (Stiglmayr, Wien-Föhrenau 1980)
32.82 F. Födermayr: *Zur gesanglichen Stimmgebung in der außereuropäischen Musik*, Vol. 1 and 2 (Stiglmayr, Wien-Föhrenau 1971)
32.83 R. Randall: *Frequency Analysis*, 3rd edn. (Bruel Kjaer, Naerum 1987)
32.84 R. McAulay, T. Quatieri: Speech analsis/synthesis based on sinusoidal representation, IEEE Trans. on Acoustics, Speech, and Signal Processing **34**, 744–754 (1986)
32.85 A. Schneider: *Tonhöhe – Skala – Klang. Akustische, tonometrische und psychoakustische Studien auf vergleichender Grundlage* (Orpheus, Bonn 1997)
32.86 A. Schneider: Inharmonic sounds: Implications as to pitch, timbre, and consonance, J. New Music Res. **29**, 275–301 (2000)
32.87 G. de Poli, A. Piccialli, C. Roads (Eds.): *Representations of Musical Signals* (MIT Press, Cambridge 1991)
32.88 C. Roads, S. Pope, A. Piccialli, G. de Poli (Eds.): *Musical Signal Processing* (Swets Zeitlinger, Lisse, Abingdon 1997)
32.89 G. Peeters, B. Giordano, P. Susini, N. Misdariis, St McAdams: The timbre toolbox: Extracting audio descriptors from musical signals, J. Acoust. Soc. Am. **130**, 2902–2916 (2011)
32.90 H. von Helmholtz: *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik* (Vieweg, Braunschweig 1863), 3rd edn. 1870, 6th edn. 1913
32.91 C. Stumpf: *Tonpsychologie*, Vol. 2 (Barth, Leipzig 1890)
32.92 C. Stumpf: *Die Sprachlaute* (Springer, Berlin 1926)
32.93 W. Köhler: Akustische Untersuchungen I–III, Z. Psychol. **54**, 241–289 (1910)
32.94 W. Köhler: Akustische Untersuchungen I–III, Z. Psychol. **58**, 59–140 (1911)
32.95 W. Köhler: Akustische Untersuchungen I–III, Z. Psychol. **64**, 241–289 (1913)
32.96 W. Köhler: Akustische Untersuchungen I–III, Z. Psychol. **72**, 1–192 (1915)
32.97 C. Stumpf: *Konsonanz und Dissonanz* (Barth, Leipzig 1898)
32.98 C. Stumpf: Beobachtungen über Kombinationstöne, Z. Psychol. **55**, 1–142 (1910)
32.99 C. Stumpf: Über neuere Untersuchungen zur Tonlehre, Beitr. Akust. Musikwiss. **8**, 305–344 (1914)
32.100 E. von Hornbostel: Psychologie der Gehörserscheinungen. In: *Handbuch der normalen und pathol. Physiol.*, Vol. 11, ed. by A. Bethe (Springer, Berlin 1926) pp. 701–730
32.101 G. Rich: A preliminary study of tonal volume, J. Exp. Psych. **1**, 13–22 (1916)
32.102 S. Stevens: The volume and intensity of tones, Am. J. Psych. **46**, 397–408 (1934)
32.103 S. Stevens: The attributes of tones, Proc. Natl. Acad. Sci. **20**, 457–459 (1934)
32.104 E. Terhardt: *Akustische Kommunikation* (Springer, Berlin 1998)
32.105 W. Lichte: Attributes of complex tones, J. Exp. Psychol. **28**, 455–480 (1941)
32.106 G. Albersheim: *Zur Psychologie der Ton- und Klangeigenschaften unter Berücksichtigung der Zweikomponententheorie und der Vokalsystematik* (Heitz, Leipzig, Straßburg 1939), repr. Körner, Baden-Baden 1975
32.107 G. Rich: A Study of tonal attributes, Am. J. Psychol. **30**, 121–164 (1919)
32.108 C. Ruckmick: A new classification of tonal qualities, Psych. Rev. **36**, 172–180 (1929)
32.109 A. Wellek: Die Mehrseitigkeit der „Tonhöhe" als Schlüssel zur Systematik der musikalischen Erscheinungen, Z. Psychol. **134**, 302–348 (1935)
32.110 H. Ebbinghaus: *Grundzüge der Psychologie*, Vol. 1, 4th edn. (Veit, Leipzig 1919)
32.111 ASA: *American Standard Acoustical Terminology* (ASA, New York 1960) p. 45
32.112 F. Kittler: *Gramophone, Film, Typewriter* (Stanford Univ. Press, Stanford 1999)
32.113 Elektronische Musik: *Sonderheft über elektronische Musik, Technische Hausmitteilungen des Nordwestdeutschen Rundfunks*, Jg. 6, Nr. 1/2 (NWDR, Cologne 1954)
32.114 A. Moles: *Théorie de l'information et perception ésthetique* (Flammarion, Paris 1958)
32.115 D. Gabor: Theory of communication, J. Inst. Electr. Eng. **93**, 429–457 (1946)
32.116 C. Shannon: A mathematical theory of communication, Bell System Techn. J. **27**(379–423), 623–656 (1948)

32.117 L. Russolo: *L'arte dei rumori* (Ed. Futuriste di Poesia, Milano 1913) French transl.: *L'art des bruits. Manifeste futuriste 1913* (Richard-Masse, Paris 1954)

32.118 P. Schaeffer: *Traité des objets musicaux. Nouveau Edition* (Seuil, Paris 1977)

32.119 E. Husserl: *Erfahrung und Urteil. Untersuchungen zur Genealogie der Logik*, 5th edn. (Meiner, Hamburg 1976), ed. by L. Landgrebe

32.120 K. Hevner: Experimental studies of the elements of expression in music, Am. J. Psychol. **48**, 246–268 (1936)

32.121 C. Osgood, G. Suci, P. Tannenbaum: *The Measurement of Meaning* (Univ. of Illinois Press, Urbana 1957)

32.122 S. Ertel: Standardisierung eines Eindrucksdifferentials, Z. exp. angew. Psychol. **12**, 22–58 (1965)

32.123 R. Kendall, E. Carterette: Verbal attributes of simultaneous wind instrument timbres: I. von Bismarck's adjectives, Music Percept. **10**, 445–468 (1993)

32.124 A. Schneider, D. Müllensiefen: Musikpsychologie in Hamburg. Ein Forschungsbericht, Syst. Musikwiss. – Syst. Musicol. **7**, 59–89 (2000)

32.125 H. Böttcher, U. Kerner: *Methoden der Musikpsychologie* (Edition Peters, Leipzig 1978)

32.126 S. Mulaik: *Foundations of Factor Analysis*, 2nd edn. (CRC, Boca Raton 2010)

32.127 R. Mores: Nasality in musical sounds – a few intermediate results. In: *Systematic Musicology: Empirical and theoretical Studies*, ed. by A. Schneider, A. von Ruschkowski (Lang, Frankfurt am Main 2011) pp. 127–136

32.128 St Solomon: Semantic approach to the perception of complex sounds, J. Acoust. Soc. Am. **30**, 421–425 (1958)

32.129 V. Rahls: *Psychometrische Untersuchungen zur Wahrnehmung musikalischer Klänge*, Ph.D. thesis (Univ. of Hamburg, Hamburg 1966)

32.130 E. Jost: *Akustische und psychometrische Untersuchungen an Klarinettenklängen* (A. Volk, Köln 1967)

32.131 G. von Bismarck: Timbre of steady sounds: A factorial investigation of its verbal attributes, Acustica **30**, 146–159 (1974)

32.132 R. Kendall, E. Carterette: Perceptual scaling of simultaneous wind instrument timbres, Music Percept. **8**, 369–404 (1991)

32.133 W. Wundt: *Grundriß der Psychologie*, 15th edn. (Engelmann, Leipzig 1928)

32.134 G. von Bismarck: Sharpness as an attribute of the timbre of steady sounds, Acustica **30**, 159–172 (1974)

32.135 L. Marks: On cross-modal similarity: the perceptual structure of pitch, loudness, and brightness, J. Exp. Psychol.: Hum. Percept. Perform. **15**, 586–602 (1989)

32.136 E. Zwicker, H. Fastl: *Psychoacoustics. Facts and Models*, 2nd edn. (Springer, Berlin 1999)

32.137 S. Stevens, J. Harris: The scaling of subjective roughness and smoothness, J. Exp. Psychol. **64**, 489–494 (1962)

32.138 A. Schneider: 'Verschmelzung', tonal fusion, and consonance: Carl Stumpf revisited. In: *Music, Gestalt, and Computing. Studies in Cognitive and Systematic Musicology*, ed. by M. Leman (Springer, Berlin 1997) pp. 117–143

32.139 Ch Seeger: *Studies in Musicology. Vol. I (1935–1975)* (Univ. of Cal. Press, Berkeley 1977)

32.140 W. Thies: *Grundlagen einer Typologie der Klänge* (Wagner, Hamburg 1982)

32.141 R. Kendall, E. Carterette: Perceptual scaling of simultaneous wind instrument timbres: II. Adjectives induced from Piston's 'Orchestration', Music Percept. **10**, 469–502 (1993)

32.142 J. Hajda: The Effect of dynamic acoustical features on musical timbre. In: *Analysis, Synthesis, and Perception of Musical Sounds*, ed. by J. Beauchamp (Springer, New York 2007) pp. 250–271

32.143 A. Zacharakis, K. Pastiadis, J. Reiss: An interlanguage study of musical timbre semantic dimensions and their acoustic correlates, Music Percept. **31**, 339–358 (2013)

32.144 A. Nykänen, Ö. Johansson, J. Lundberg, J. Berg: Modelling perceptual dimensions of saxophone sounds, Acustica **95**, 539–549 (2009)

32.145 W. Aures: Der sensorische Wohlklang als Funktion psychoakustischer Empfindungsgrößen, Acustica **58**, 282–290 (1985)

32.146 W. Aures: Berechnungsverfahren für den sensorischen Wohlklang beliebiger Schallsignale, Acustica **59**, 130–141 (1985)

32.147 A. Tversky: Features of similarity, Psych. Rev. **84**, 327–352 (1977)

32.148 D. Medin, R. Goldstone, D. Gentner: Respects for similarity, Psych. Rev. **100**, 254–278 (1993)

32.149 C. Stumpf: *Tonpsychologie*, Vol. 1 (Barth, Leipzig 1883)

32.150 W. Torgerson: *Theory and Method of Scaling* (Wiley, New York 1958)

32.151 F. Sixtl: *Meßmethoden der Psychologie. Theoretische Grundlagen und Probleme*, 2nd edn. (Beltz, Weinheim, Basel 1982)

32.152 I. Borg, J. Lingoes: *Multidimensional Similarity Structure Analysis* (Springer, New York 1987)

32.153 I. Borg, P. Groenen: *Modern Multidimensional Scaling. Theory and Applications*, 2nd edn. (Springer, New York 2005)

32.154 F.G. Ashby, N. Perrin: Toward a unified theory of similarity and recognition, Psychol. Rev. **95**, 124–150 (1988)

32.155 N. Perrin: Uniting identification, similarity and preference: General recognition theory. In: *Multidimensional Models of Perception and Cognition*, ed. by F. Ashby (Erlbaum, Hillsdale 1992) pp. 123–145

32.156 R. Nosofsky: Similarity scaling and cognitive process models, Ann. Rev. Psych. **43**, 25–53 (1992)

32.157 J. Beran: *Statistics in Musicology* (Chapman Hall, Boca Raton 2004)

32.158 S. Donnadieu: Mental representations of the timbre of complex sounds. In: *Analysis, Synthesis, and Perception of Musical Sounds*, ed. by

J. Beauchamp (Springer, New York 2007) pp. 272–319
32.159 J. Grey: An Exploration of Musical Timbre. Report no. Stan-M-2 (CCRMA Dept. of Music, Stanford 1975)
32.160 J. Grey: Multidimensional perceptual scaling of musical timbres, J. Acoust. Soc. Am. **61**, 1270–1277 (1977)
32.161 J. Grey, J. Gordon: Perceptual effects of spectral modifications on musical timbres, J. Acoust. Soc. Am. **63**, 1493–1500 (1978)
32.162 R. Plomp, H. Steeneken: Effect of phase on the timbre of complex tones, J. Acoust. Soc. Am. **46**, 409–421 (1969)
32.163 J. Miller, E. Carterette: Perceptual space for musical structures, J. Acoust. Soc. Am. **58**, 711–720 (1975)
32.164 D. Wessel: Timbre space as musical control structure, Comp. Music J. **3**, 45–52 (1979)
32.165 C. Krumhansl: Why is musical timbre so hard to understand? In: *Structure and Perception of Electroacoustic Sound and Music*, ed. by S. Nielzen, O. Olsson (Elsevier, Amsterdam 1989) pp. 43–53
32.166 P. Iverson, C. Krumhansl: Isolating the dynamic attributes of musical timbre, J. Acoust. Soc. Am. **94**, 2595–2603 (1993)
32.167 St McAdams, S. Winsberg, S. Donnadieu, G. de Soete, J. Krimphoff: Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes, Psychol. Res. **58**, 177–192 (1995)
32.168 B. Markuse, A. Schneider: Ähnlichkeit, Nähe, Distanz: zur Anwendung multidimensionaler Skalierung in musikwissenschaftlichen Untersuchungen, Syst. Musikwiss. – Syst. Musicol. **4**, 53–89 (1996)
32.169 R. Plomp: *Aspects of Tone Sensation* (Academic, London 1976)
32.170 A. Caclin, St McAdams, B. Smith, S. Winsberg: Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones, J. Acoust. Soc. Am. **118**, 471–482 (2005)
32.171 S. Dixon: Onset detection revisited. In: *Proc. 9th Intern. Conf. Digital Audio Effects (DAFx-06), Montreal* (2006) pp. 133–137
32.172 W.R. Garner: *The Processing of Information and Structure* (Erlbaum, Potomac 1974)
32.173 B. Kostek: *Perception-Based Data Processing in Acoustics. Applications to Music Information Retrieval and Psychophysiology of Hearing* (Springer, Berlin 2005)
32.174 St Handel: Timbre perception and auditory object identification. In: *Hearing*, ed. by B. Moore (Academic, San Diego 1995) pp. 425–461
32.175 J. Schouten: The perception of timbre. In: *Reports 6th Int. Congr. Acoust., Tokyo*, Vol. VI (1968) pp. 35–44
32.176 J. Licklider: Basic correlates of the auditory stimulus. In: *Handbook of Experimental Psychology*, ed. by S.S. Stevens (Wiley, New York 1951) pp. 985–1039
32.177 W. Slawson: *Sound Color* (Univ. of Cal. Press, Berkeley 1985)
32.178 H. Pollard, E. Jansson: Analysis and assessment of musical starting transients, Acustica **51**, 249–262 (1982)
32.179 D. Howard, J. Angus: *Acoustics and Psychoacoustics*, 2nd edn. (Focal, Oxford 2001)
32.180 E. Jost: Über den Einfluß der Darbietungsdauer auf die Identifikation von instrumentalen Klangfarben. In: *Jahrb. Staatl. Inst. Musikforsch. (Berlin) für 1969* (1970) pp. 83–92
32.181 C. Reuter: *Die auditive Diskrimination von Orchesterinstrumenten. Verschmelzung und Heraushörbarkeit von Instrumentalklangfarben im Ensemblespiel* (Lang, Frankfurt am Main 1996)
32.182 R. Rasch: The Perception of simultaneous notes such as in polyphonic music, Acustica **40**, 21–33 (1978)
32.183 P. Vos, R. Rasch: The perceptual onset of tones, Percept. Psychophys. **29**, 323–335 (1981)
32.184 B. Lau, R. Bader, A. Schneider, P. Wriggers: Finite Element transient calculation of a bell struck by its clapper. In: *Concepts, Experiments, and Fieldwork: Studies in Systematic Musicology and Ethnomusicology*, ed. by R. Bader, C. Neuhaus, U. Morgenstern (Lang, Frankfurt am Main 2010) pp. 137–156
32.185 A. Schneider, M. Leman: Sonological and psychoacoustic characteristics of carillon bells. In: *The Quality of Bells: Proc. of the 16th Meeting of the FWO Res. Soc. Foundations Music Research*, ed. by M. Leman (Univ. of Ghent., Ghent 2002)
32.186 A. Schneider, M. Leman: Sound, pitches and tuning of a historic carillon. In: *Studies in Musical Acoustics and Psychoacoustics*, ed. by A. Schneider (Springer, Cham 2017) pp. 247–298
32.187 A. Melka: Messungen der Klangeinsatzdauern bei Musikinstrumenten, Acustica **23**, 108–177 (1970)
32.188 T. Gäumann: The pretransient of the harpsichord sound. In: *Proc. Stockholm Musical Acoustics Conf. (SMAC '03)*, Vol. I (2003) pp. 163–166
32.189 A. Beurmann, A. Schneider: Sonological analysis of harpsichord sounds. In: *Proc. Stockholm Music Acoustics Conf. (SMAC '03)*, Vol. I (2003) pp. 167–170
32.190 E. Saldanha, J. Corso: Timbre cues and identification of musical instruments, J. Acoust. Soc. Am. **36**, 2021–2026 (1964)
32.191 L. Wedin, G. Goude: Dimension analysis of the perception of instrumental timbres, Scand. J. Psych. **13**, 228–240 (1972)
32.192 N. Fletcher: Acoustical correlates of flute performance technique, J. Acoust. Soc. Am. **57**, 233–237 (1975)
32.193 X. Serra: Musical sound modelling with sinusoids plus noise. In: *Musical Signal Processing*, ed. by C. Roads, S. Pope, A. Piccialli, G. de Poli (Swets Zeitlinger, Lisse 1997) pp. 91–122
32.194 S. Levine, J. Smith III: A compact and malleable sines + transients + noise model for sound. In: *Analysis, Synthesis, and Perception of Musical Sound*, ed. by J. Beauchamp (Springer, New York 2007) pp. 145–174

32.195 R. Patterson, E. Gaudrain, Th Walters: The Perception of family and register in musical tones. In: *Music Perception*, ed. by M. Riess Jones, R. Fay, A. Popper (Springer, New York 2010) pp. 13–50

32.196 W. Adelung: *Einführung in den Orgelbau*, 3rd edn. (Breitkopf Härtel, Leipzig 1972)

32.197 N. Fletcher, T. Rossing: *The Physics of Musical Instruments* (AIP/Springer, New York 1991)

32.198 P. Goad, D. Keefe: Timbre discrimination of musical instruments in a concert hall, Music Percept. **10**, 43–62 (1992)

32.199 J. Marozeau, A. de Cheveigné, St McAdams, S. Winsberg: The dependency of timbre on fundamental frequency, J. Acoust. Soc. Am. **114**, 2948–2957 (2003)

32.200 St Handel, M. Erickson: A rule of thumb: The bandwidth for timbre invariance is one octave, Music Percept. **19**, 121–126 (2001)

32.201 K. Steele, A. Williams: Is the bandwidth for timbre invariance only one octave?, Music Percept. **23**, 215–220 (2006)

32.202 J. Meyer: Zur klanglichen Wirkung des Streicher-Vibratos, Acustica **76**, 283–291 (1992)

32.203 J. Gärtner: *Das Vibrato unter besonderer Berücksichtigung der Verhältnisse bei Flötisten*, 2nd edn. (Bosse, Regensburg 1980)

32.204 A. Schneider, R. Mores: Fourier-Time-Transformation (FTT), analysis of sound and auditory perception. In: *Sound – Perception – Performance*, ed. by R. Bader (Springer, Cham 2013) pp. 299–329

32.205 M. Barthet, Ph Depalle, R. Kronland-Martinet, S. Ystad: Acoustical correlates of timbre and expressiveness in clarinet performance, Music Percept. **28**, 135–153 (2010)

32.206 H. Pollard, E. Jansson: A tristimulus method for the specification of musical timbre, Acustica **51**, 162–171 (1982)

32.207 E. Zwicker: *Psychoakustik* (Springer, Berlin 1982)

32.208 K. Denbigh: *Three Concepts of Time* (Springer, Berlin 1981)

32.209 A. Schneider: Virtual pitch and musical instrument acoustics. The case of idiophones. In: *Musik im virtuellen Raum. KlangArt-Kongress 1997*, ed. by B. Enders, J. Stange-Elbe (Universitätsverlag Rasch, Osnabrück 2000) pp. 397–417

32.210 A. Schneider: Sound, pitch, and scale: From 'tone measurements' to sonological analysis in ethnomusicology, Ethnomusicology **45**, 489–519 (2001)

32.211 R. Melara, L. Marks: Interaction among auditory dimensions: Timbre, pitch, and loudness, Percept. Psychophys. **48**, 169–178 (1990)

32.212 C. Krumhansl, P. Iverson: Perceptual Interactions between musical pitch and timbre, J. Exp. Psych.: Human Percept. Perform. **18**, 739–751 (1992)

32.213 E. Allen, A. Oxenham: Symmetric interactions and interferences between pitch and timbre, J. Acoust. Soc. Am. **135**, 1371–1379 (2014)

32.214 R. Patterson, R. Milroy, M. Allerhand: What is the octave of a harmonically rich note?, Contemp. Music Rev. **9**, 69–81 (1993)

32.215 H.P. Hesse: Experimente zum musikalischen Intervallurteil. In: *Jahrb. Staatl. Inst. Musikforsch. (Berlin) für 1978* (1979) pp. 72–87

32.216 M. van Tongeren: *Overtone Singing. Physics and Metaphysics of Harmonics in East and West* (Fusica, Amsterdam 2002)

32.217 St Handel: *Listening. An Introduction to the Perception of Auditory Events* (MIT Press, Cambridge 1989)

32.218 A. Schneider, K. Frieler: Perception of harmonic and inharmonic sounds: Results from ear models. In: *Computer Music Modeling and Retrieval. Genesis of Meaning in Sound and Music*, ed. by S. Ystad, R. Kronland-Martinet, K. Jensen (Springer, Berlin 2009) pp. 18–44

32.219 J. Chowning: John Chowning on composition. In: *Composers and the Computer*, ed. by C. Roads (W. Kaufmann, Los Altos 1985) pp. 18–25

32.220 J. Chowning: *Phoné (1979–80)*, CD (Wergo/Schott, Mainz 1988)

32.221 D. Arfib, F. Keiler, U. Zölzer: Source-Filter Processing. In: *DAFX. Digital Audio Effects*, ed. by U. Zölzer (Wiley, Chichester 2002) pp. 299–372

32.222 J. Keuler: Problems of shape and background in sounds with inharmonic spectra. In: *Music, Gestalt, and Computing. Studies in Cognitive and Systematic Musicology*, ed. by M. Leman (Springer, Berlin 1997) pp. 214–224

32.223 F. Lerdahl: Timbral hierarchies, Contemp. Music Rev. **2**, 135–160 (1987)

32.224 A. Schönberg: *Harmonielehre*, 2nd edn. (Universal Ed., Wien 1922)

32.225 A. Schneider: Akustische und psychoakustische Anmerkungen zu Arnold Schönbergs Emanzipation der Dissonanz und zu seiner Idee der Klangfarbenmelodie, Hamburger Jahrb. Musikwiss. **17**, 35–55 (2000)

32.226 A. Fokker: *New Music with 31 Notes* (Verlag für Systematische Musikwissenschaft, Bonn 1975)

32.227 W. Voigt: *Dissonanz und Klangfarbe. Instrumentationsgeschichtliche und experimentelle Untersuchungen* (Verlag für Systematische Musikwissenschaft, Bonn 1985)

32.228 P. Boersma, G. Kovacic: Spectral characteristics of three styles of Croatian folk singing, J. Acoust. Soc. Am. **119**, 1805–1816 (2006)

32.229 A. Bregman: *Auditory Scene Analysis* (MIT Press, Cambridge 1990)

32.230 Y. Ando: *Concert Hall Acoustics* (Springer, Berlin 1985)

32.231 L. Beranek: *Concert and Opera Halls. How they sound* (Acoust. Soc. Am., Woodbury 1996)

32.232 H. Kuttruff: *Room Acoustics*, 5th edn. (Spon, London 2009)

32.233 J. Blauert: *Spatial Hearing. The Psychophysics of Human Sound Localization*, 6th edn. (MIT Press, Cambridge 2008)

32.234 Ch Brown, B. May: Comparative mammalian sound localization. In: *Sound Source Localization*, ed. by A. Popper, R. Fay (Springer, New York 2005) pp. 124–178

32.235 K. Hancock, B. Delgutte: A physiologically based model of interaural time difference discrimination, J. Neurosci. **24**, 7110–7117 (2004)

32.236 F. Wightman, D. Kistler: Sound localization. In: *Human Psychophysics*, ed. by W. Yost, A. Popper, R. Fay (Springer, New York 1993) pp. 155–192

32.237 T. Ziemer: Psychoacoustic effects in wave field synthesis applications. In: *Systematic Musicology: Empirical and Theoretical Studies*, ed. by A. Schneider, A. von Ruschkowski (Lang, Frankfurt am Main 2011) pp. 153–162

32.238 J. Eggermont: Between sound and perception: Reviewing the search for a neural code, Hearing Res. **157**, 1–42 (2001)

32.239 I. Nelken: Processing of complex stimuli and natural scenes in the auditory cortex, Curr. Opin. Neurobiol. **14**, 474–480 (2004)

32.240 I. Nelken: Processing of complex sounds in the auditory system, Curr. Opin. Neurobiol. **18**, 413–417 (2008)

32.241 C. Von der Malsburg: Binding in models of perception and brain function, Curr. Opin. Neurobiol. **5**, 520–526 (1995)

32.242 A. Roskies: The binding problem, Neuron **24**, 7–9 (1999)

32.243 B. Moore: Frequency analysis and pitch perception. In: *Human Psychophysics*, ed. by W. Yost, A. Popper, R. Fay (Springer, New York 1993) pp. 56–115

32.244 B. Moore: Frequency analysis and masking. In: *Hearing*, ed. by B. Moore (Academic, San Diego 1995) pp. 161–205

32.245 B. Moore: Basic psychophysics of human spectral processing. In: *Auditory Spectral Processing*, International Review of Neurobiology, Vol. 70, ed. by M. Malmierca, D. Irvine (Elsevier, Amsterdam 2005) pp. 49–86

32.246 E. Terhardt, G. Stoll, M. Seewann: Algorithm for extraction of pitch and pitch salience from complex tone signals, J. Acoust. Soc. Am. **71**, 679–688 (1982)

32.247 E. Terhardt, G. Stoll, M. Seewann: Pitch of complex signals according to virtual-pitch theory: Tests, examples, and predictions, J. Acoust. Soc. Am. **71**, 671–678 (1982)

32.248 E. Terhardt, M. Seewann: Auditive und objektive Bestimmung der Schlagtonhöhe von historischen Kirchenglocken, Acustica **54**, 129–144 (1984)

32.249 D. Huron: Voice denumerability in polyphonic music of homogeneous timbres, Music Percept. **6**, 361–382 (1989)

32.250 R. Rasch: Synchronization in performed ensemble music, Acustica **43**, 121–131 (1979)

32.251 St McAdams: Spectral fusion and the creation of auditory images. In: *Music, Mind, and Brain*, ed. by M. Clynes (Plenum, London 1982) pp. 279–298

32.252 St McAdams: Segregration of concurrent sounds. I: Effects of frequency modulation coherence, J. Acoust. Soc. Am. **86**, 2148–2159 (1989)

32.253 D. Mellinger, B. Mont-Reynaud: Scene analysis. In: *Auditory Computation*, ed. by H. Hawkins, T. McMullen, A. Popper, R. Fay (Springer, New York 1996) pp. 271–331

32.254 J. Van Eyck: Der Fluyten Lust-Hof, beplant met Psalmen, Pavanen, Almanden, Couranten, Balletten, Airs (Matthysz, Amsterdam 1654)

32.255 L. Van Noorden: *Temporal Coherence in the Perception of Tone Sequences*, Ph.D. Thesis (Technical Univ. of Eindhoven, Eindhoven 1975)

32.256 G. Kubik: Die Amadinda-Musik von Buganda. In: *Musik in Afrika*, ed. by A. Simon (Museum für Völkerkunde, Berlin 1983) pp. 139–165

32.257 G. Kubik: *Theory of African Music*, Vol. I (Heinrichshofen, Wilhelmshaven 1994)

32.258 U. Wegner: Cognitive aspects of Amadinda xylophone music from Buganda: Inherent patterns reconsidered, Ethnomusicology **37**, 201–241 (1993)

32.259 G. Miller, G. Heise: The trill threshold, J. Acoust. Soc. Am. **22**, 637–638 (1950)

32.260 A. Bregman, J. Campbell: Primary auditory stream segregation and perception of order in rapid sequences of tones, J. Exp. Psych. **89**, 244–249 (1971)

32.261 P. Iverson: Auditory stream segregation by musical timbre. Effects of static and dynamic acoustic attributes, J. Exp. Psych.: Hum. Percept. Perf. **21**, 751–763 (1995)

32.262 B. Roberts, B. Glasberg, B. Moore: Primitive stream segregation of tone sequences without differences in fundamental frequency or passband, J. Acoust. Soc. Am. **112**, 2074–2085 (2002)

32.263 H. Helson: The fundamental propositions of Gestalt Psychology, Psych. Rev. **40**, 13–32 (1933)

32.264 A. Fischer: *Neurophysiologisch motivierte Modelle zur akustischen Figur-Hintergrund-Trennung* (Deutsch, Frankfurt am Main 1994)

32.265 S. McCabe, M. Denham: A model of auditory streaming, J. Acoust. Soc. Am. **101**, 1611–1621 (1997)

32.266 A. Klapuri: Auditory model-based methods for multiple fundamental frequency estimation. In: *Signal Processing Methods for Music Transcription*, ed. by A. Klapuri, M. Davy (Springer, New York 2006) pp. 229–266

32.267 A. Klapuri: Multipitch analysis of polyphonic music and speech signals using an auditory model, IEEE Transactions Audio, Speech, and Language Process. **16**, 255–266 (2008)

32.268 K. Kashino: Auditory scene analysis in music signals. In: *Signal Processing Methods for Music Transcription*, ed. by A. Klapuri, M. Davy (Springer, New York 2006) pp. 299–325

32.269 M. Goto: Music scene description. In: *Signal Processing Methods for Music Transcription*, ed. by A. Klapuri, M. Davy (Springer, New York 2006) pp. 327–359

32.270 F. Cañadas Quesada, N. Ruiz Reyes, P.V. Candeas, J. Carabias, S. Maldonado: A multiple-F0 estimation approach based on Gaussian spectral modelling for polyphonic music transcription, J. New Music Res. **39**, 93–107 (2010)

32.271 A. Schneider: Klanganalyse als Methodik der Popularmusikforschung, Hamburger Jahrb. Musikwiss. **19**, 107–129 (2002)