Perception of Harmonic and Inharmonic Sounds: Results from Ear Models

Albrecht Schneider and Klaus Frieler

Institute for Musicology, University of Hamburg, Germany,
{aschneid, klaus.frieler}@uni-hamburg.de

Abstract. We report on experiments in which musically relevant harmonic and inharmonic sounds have been fed into computer-based ear models (or into modules which at least simulate parts of the peripheral auditory system) working either in the frequency or in the time domain. For a major chord in just intonation, all algorithms produced reliable and interpretable output, which explains mechanisms of pitch perception. One model also yields data suited to demonstrate how sensory consonance and 'fusion' are contained in the ACF of the neural activity pattern.
With musical sounds from instruments (carillon, *gamelan*) which represent different degrees of inharmonicity, the performance of the modules reflects difficulties in finding correct spectral and/or virtual pitch(es) known also from behavioral experiments. Our measurements corroborate findings from neurophysiology according to which much of the neural processing relevant for perception of pitch and consonance is achieved subcortically.

1 Introduction

During the past decades, a vast amount of research in sensation and perception of sounds has been undertaken in both sensory physiology and psychophysics, respectively (e.g., Popper & Fay 1992, Ehret & Romand 1997, Terhardt 1998, Zwicker & Fastl 1999, Plack et al. 2005). At the same time, the field of music perception gained new impetus due to approaches influenced by cognitive psychology (e.g., Sloboda 1985, Krumhansl 1990, Bregman 1990), or by cognitive science in general (e.g., Balaban et al. 1992). There have been efforts to bring together facts and models from both fields (e.g., Handel 1989, Bregman 1990, McAdams & Bigand 1993, Leman 1995), however, many problems still wait to be investigated.

In the following, we shall deal with the perception of inharmonic sounds as they are found in a number of music cultures. One reason to do so is that most of the experiments in hearing, and in sound perception in general, have been conducted with periodic sounds having harmonic spectra, or with inharmonic sounds which have little if any relevance for music (e.g., white or pink noise). Another reason is that perception of inharmonic sounds can be simulated with computer-based ear models in a bottom-up approach whereby certain mechanisms as well as limits of perceptual analysis might become evident. These, in turn, could have implications for music cognition. As will be demonstrated in this chapter by experimental data obtained from measurements done with computer ear models, our auditory system constrains perception of sound stimuli and thereby also influences cognition.

2 Perception of Harmonic Sounds

Empirical research in auditory perception almost always has a focus on pitch since it is a fundamental property of sounds (Houtsma 1995, Terhardt 1998, Plack & Oxenham 2005), which may be used in various systems of human communication including speech and music. Typically, explorations of pitch stem from periodic time signals such as sinusoids or complex tones comprising a number of harmonics in addition to a fundamental frequency. It has been found in many experiments that the pitch perceived from such periodic signals corresponds closely to the fundamental frequency of a complex harmonic tone, or to the frequency with which a complex waveshape composed of harmonics lacking a fundamental repeats per time unit. Different from abbreviations common in psychoacoustics, we will label the fundamental f_1 (since it is the lowest partial of a harmonic spectrum), and the repetition frequency of a harmonic complex lacking a fundamental, f_0. Fig. 1 shows the time function $y(t)$ of a harmonic complex tone (10 partials) with f_1 at 100 Hz, and partial amplitudes $A_n = 1/n$ (n = harmonic number 1, 2, 3,. . .). The length of the period according to $T = 1/f$ is 10 ms. Fig. 2 shows the same signal of which partials 1-3 have been removed while f_0 has been added to the graph as an extra (sinusoidal) component ($f_0 = 100\,\mathrm{Hz}, A = 1$) to indicate that the two signals are likely to yield the same pitch in subjects. This model implies that our system of hearing includes a mechanism for periodicity extraction from sound signals such as speech and music. Periodicity extraction has been a major issue in hearing theory since long (cf. de Hesse 1972, de Boer 1976, Lyon & Shamma 1996, Schneider 1997a/b, 2000a, de Cheveigné 2005, Langner 2007).

The steep wavecrests at the onset of each vibration period as obvious from Fig. 1 (and also Fig. 2), in accordance with the 'volley principle' of Wever and Bray (Wever 1949, chs. 8, 9) have been regarded as triggering synchronized trains of neural spikes, which are suited to elicit a stable pitch percept. The same pitch, though is heard if the phase relations between signal components are changed so that no strong peaks are found at the onset of each period yet a certain periodicity of the signal is retained (Schneider 1997b, 123-135). The change of phase relations can affect the salience of pitch, and will often result in a change of the timbral quality of a given sound.

Fig. 1. Harmonic complex tone, 10 partials, $f_1 = 100$ Hz, $A_n = 1/n$

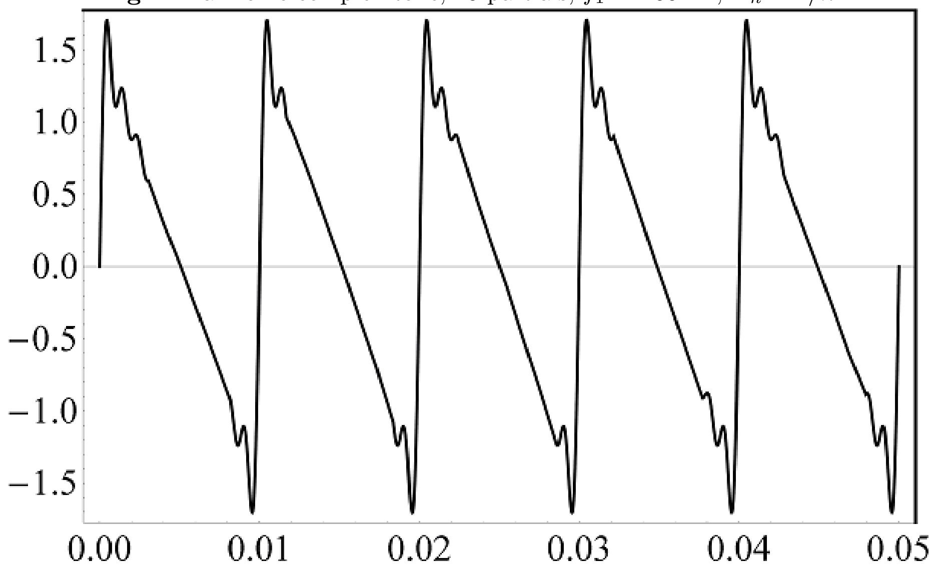
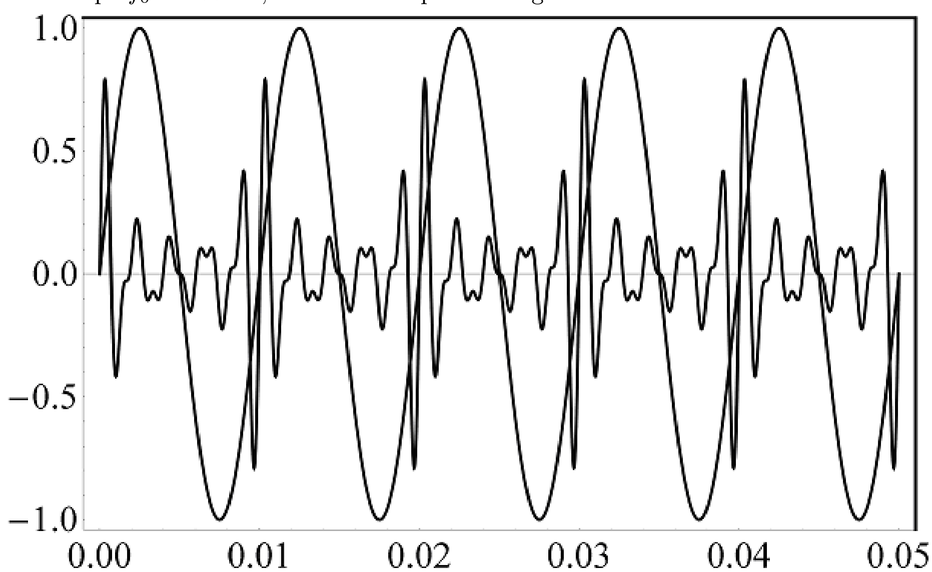


Fig. 2. Harmonic complex tone, partials 4-10, plus repetition frequency of complex waveshape $f_0 = 100$ Hz, A = 1 of the periodic signal



3 Harmonicity, Consonance and Fusion (*Verschmelzung*)

It is possible to construct very stable chords from complex tones which have a number of harmonics being phase-locked. For example, if one builds a chord from three complex tones having fundamental frequencies $\{300\,\text{Hz}, 400\,\text{Hz}, 500\,\text{Hz}\}$ and seven harmonics each with amplitudes defined by $A = 1/n$ $(n = 1, 2, \ldots)$, the resulting sound represents a major chord in just intonation (Schneider 1997a, Fig. 1). Such chords which in real instruments are available from organ mixture stops, had been used by the psychologist Carl Stumpf to study the phenomenon of *Verschmelzung*, which can be regarded as a perceptual and cognitive quality experienced when listening attentively to chords such as the example given here. Due to the strict periodicity of the waveshape of the respective sound as well as to the likewise perfect harmonicity of the spectral components making up the complex tones, the sound offers an optimum of 'fusion', on the one hand, and still allows identification of many partials, on the other. Moreover, due to the strict periodicity of the sound, one perceives a low pitch at 100 Hz, which corresponds to the f_0. One can regard the f_0 as a virtual *basse fondamentale* whose function as the base note of major chords was explained in Rameau's theory of harmony (Rameau 1722, 1737). The output of the SPINET model (Cohen et al. 1995) centered in the frequency domain alternates between 100 Hz (the repetition frequency of the complex waveshape) and 300 Hz (the lowest spectral component). With a pitch extraction algorithm operating in the time domain on a normalized autocorrelation function (ACF; Boersma 1993), the pitch assigned to the overall major chord is 100 Hz. If fed into a model of the auditory periphery (AMS; Meddis & O'Mard 1997, 2003; see below), the output is a sum ACF (SACF, Fig. 3), which aggregates periodicities found in the neural activity patterns (NAP) within the channels defined by basilar membrane (BM) filters. The aggregation across channels for the pure major chord yields strong peaks at the time lags τ (ms) listed in Tab. 1 together with the relative height of peaks expressing the AC coefficient $r_{xx'}$, and the frequencies corresponding to a certain lag $(f = 1/\tau)$. The SACF was calculated for 100 ms and is displayed in Fig. 3.

In the first column, the period T of the complex waveshape as well as its multiples are found; consequent to $f_0 = 1/T$, $f_0 = 100$ Hz is determined by the strongest peaks marking each period. Since the periods repeat identically, and $f_n = 1/n\tau$, the respective frequency values must be subharmonics $(1/2, 1/3, 1/4, \ldots)$ of f_0.

In column 4 of Tab. 1, a period corresponding to $2f_0$ as well as the fundamental frequencies of the three harmonic tones making up the major chord appear. Neglecting small numerical deviations from ideal frequency ratios, a complete harmonic series 1:2:3:4:5 (plus some periodicities representing divisions or multiples of either spectral or virtual pitches) is embedded in each period of $T =$ 10 ms. Thereby a very high degree of harmonicity is encoded in the SACF, which (provided the model is valid in regard to physiological functions) will evoke strong sensations of consonance in subjects. Moreover, sounds as that used in this experiment will also give rise to difference and combination tones if presented with sufficient SPL.

Fig. 3. SACF (100 ms), pure major chord

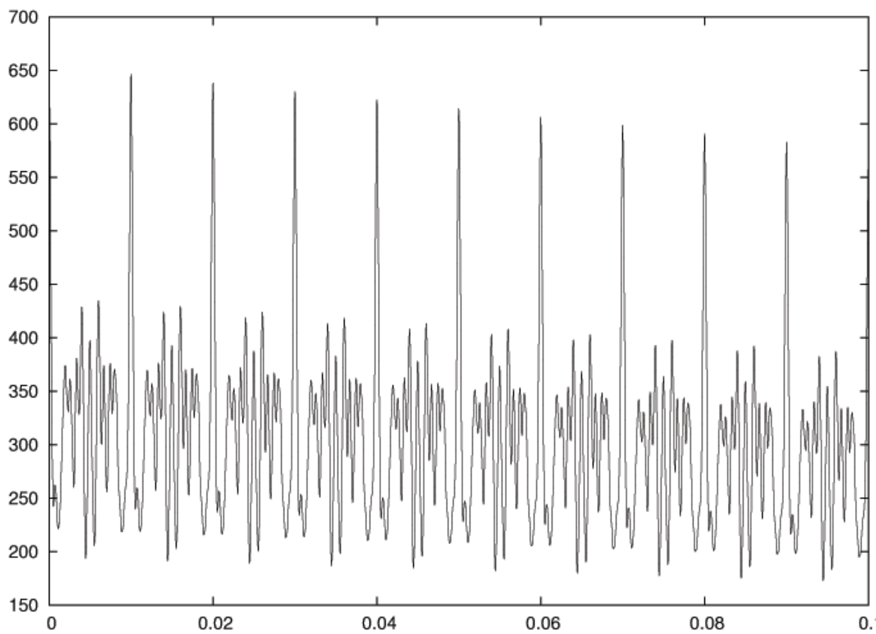


Table 1. Pure major chord, SACF; lags τ (ms), relative amplitudes of the ACF $r_{xx'}$, frequencies

τ (ms)	$r_{xx'}$	f (Hz)	τ (ms)	$r_{xx'}$	f (Hz)
10	646.63	100.00	6.02	434.76	166.09
20	638.52	50.00	3.98	428.54	251.31
30	630.44	33.33	4.98	397.72	200.84
40	622.46	25.00	3.35	380.90	298.14
50	614.49	20.00	7.46	375.83	134.08
60	606.52	16.66	6.67	374.22	150.00
70	598.62	14.29	1.98	374.01	505.26
80	590.77	12.50	8.00	370.46	125.00
90	582.94	11.11	2.54	361.11	393.44

One can take sounds such as that analyzed here as a paradigm for Stumpf's concept of consonance, which comprises both 'fusion' (*Verschmelzung*) and apperception of tonal relations (Schneider 1997a, 2008, 30-33). A musically trained listener (as was Stumpf himself), exposed to a complex harmonic chord audible for a sufficient time (> 2 seconds), can be expected to switching back and forth between a more holistic and integrative mode of hearing as well as a more analytical one; whereas the former will support the experience of 'fusion' as a Gestalt phenomenon, the latter is needed to apprehend the tonal relations between complex tones as well as between their constituents (Stumpf 1926, ch. 11).

Stumpf (1890, 1926) had already assumed a neural basis for the sensation of fusion. Since then a range of experimental data and hypotheses has been brought forward in favour of a neural basis of sensory consonance (e.g., Hesse 1972, Keidel 1989, 1992, Tramo et al. 2001, Langner 2007). Most of the approaches are based in the time domain and operate on periodicity detection in one way or another (e.g., coincidence detection of spike trains which form harmonic ratios). Some models include both temporal and spectral features.

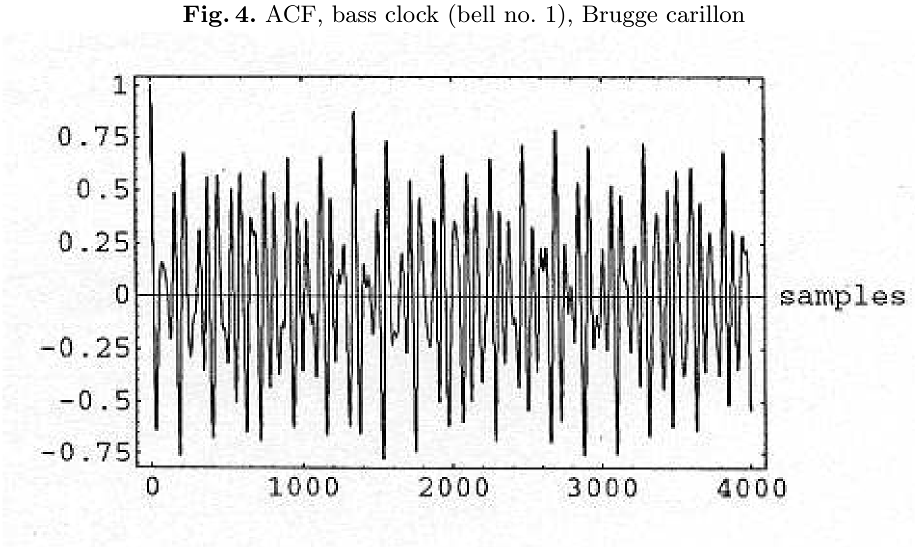
4 Perception of inharmonic sounds

Investigation of inharmonic sounds in psychoacoustics often has been pursued by either adding a constant frequency value k to all harmonic frequencies, thereby turning them into an inharmonic series, like, for example, 230, 430, 630, 830, . . . , 1630 Hz (with the original $f_1 = 200$ Hz, and $k = 30$ Hz), or by a modulation technique which yields inharmonic signals where no fundamental is present (e.g., 1230, 1430, 1630, 1830 Hz, . . . ; see de Boer 1976). In both cases, the degree of inharmonicity can be varied continuously according to the size of the constant, k. Further, a 'pseudo-fundamental' can be fitted to an incomplete series of detuned partials as long as the degree of inharmonicity is rather small (see Schneider 2000a). Such a 'pseudo-fundamental' again represents the quasi-periodicity of the signal, and hence equals f_0. With increasing inharmonicity of the spectrum, the periodicity of the signal decreases. A proven method to measure the periodicity of whatever time function $y(t)$ is the ACF. It was originally applied to the analysis of continuous movements within turbulent media as well as to the detection of periodicities in brain waves from which then a Fourier spectrum could be derived (Wiener 1961). Consequent to the Wiener-Khintchine theorem (Hartmann 1998, ch. 14), which relates the Fourier transform of the ACF of a signal to its energy spectral density, one can expect the ACF to degenerate in proportion to increasing spectral inharmonicity of a time signal $y(t)$. That is, the ACF will be the more irregular (with rather small and diminishing peaks and no clear periodicity) the more the spectral composition of the signal is shifting into inharmonicity.

Whereas the ACF of a perfectly periodic signal mirrors the periodicity $y(t) = y(t+T)$, (T = period length in samples or ms) inherent in the signal by marking prominent peaks (the correlation coefficient $r_{xx'}$ is unity at these points), the

ACF of sounds recorded from idiophones such as Western swinging or carillon bells is not just as regular. For example, the sound recorded from the bass clock of the famous carillon of Brugge (built in 1742-48 by Joris Du Mery) yields the ACF shown in Fig. 4. Though the spectrum of this bell sound contains a considerable number of quasi-harmonic components (Schneider & Leman 2002, Tab. 2 and Fig. 2), they do not all correspond to a single harmonic series in which harmonic frequencies f_n are defined by $f_n = nf_1$. Furthermore, there are quite many spectral components with inharmonic frequency ratios relative to the hum note, which in this bell is ~97.3 Hz. Consequently, the ACF of this sound, though still containing a number of peaks at certain lag points, does not exhibit a clear periodicity which could be interpreted as corresponding to the f_0 of the signal. The leftmost strong peak after onset (for which $r_{xx'} = 1$) occurs after a lag of τ ~1350 samples, which, for a sampling frequency of 44.1 kHz, corresponds to ~30 ms and yields a frequency of, roughly, 33.3 Hz, which is about 1/3 of the lowest spectal component contained in the bell sound.

Fig. 4. ACF, bass clock (bell no. 1), Brugge carillon



Sounds from many other idiophones such as gong chimes found in Javanese and Balinese *gamelan*, respectively, are much more inharmonic in spectral composition than is the carillon bell we have refered to (for examples and detailed analyses, see Schneider 1997b). In extreme cases such as shallow gongs like the Chinese *tam-tam*, the spectrum is very inharmonic, and also very dense with spectral components which interact, giving rise to amplitude modulation (AM). In addition, due to certain nonlinearities in the pattern of vibration, modal frequencies can be quite unstable so that the whole sound becomes transitory and fluctuating. In the *tam-tam*, the ACF drops to $r_{xx'}$ ∼0.5 or less immediately after onset of the sound, and then goes down to $r_{xx'}$ ∼0.25 (Schneider & Bader 2003, Figs. 5 and 6), which indicates that there is very little temporal and spectral coherence in the sound. As a result of both the complex pattern of vibration, and the likewise complex inharmonic spectrum, no definite pitch can be assigned to such a sound.

In regard to sounds and the pitch they may evoke, one can distinguish two extremes which may be considered as marking the two poles of dimensions suited to classify sounds in regard to pitch perception and pitch salience (see below):

1. A sinusoidal of given frequency and amplitude, which, as a stationary signal, yields one stable pitch percept corresponding to the frequency of the signal $y(t) = A\sin(2\pi ft)$.
2. A complex inharmonic sound which comprises many spectral components irregularly spaced along the frequency axis, and which undergoes AM (as well as some FM since modal frequencies can be unstable). The *tam-tam* is a clear example of this type of sounds which lack periodicity and do not yield any clear pitch percept. Rather, such sounds have a sound colour or timbral quality (in this case, a metallic, clangy and swirling sound quality) which may indicate a certain pitch area on a broad scale from low to high, yet do not give rise to any definite pitch. As is known from experiments with filtered noise bands, even these can be arranged to form a rough sequence of 'pitches' or, rather, different degrees of brightness (Hesse 1982), which, due to the interaction of the variables of tone height and tonal brightness, *appear* as 'pitches'.

The dimensions (quasi-continua) can be arranged according to the degree of periodicity of the time signal as well as the harmonicity of the corresponding spectrum; in addition, the psychoacoustic attributes of pitch and pitch salience (cf. Terhardt 1998, ch. 11) can be matched to the other features. Without going into details of signals and systems theory (Bachmann 1992, Terhardt 1998, Hartmann 1998), and allowing for some simplifications, we can establish the following bipartite scheme:

Signal:	completely/predominantly stationary	predominantly transient
Periodicity:	clear	uncertain
Spectrum:	harmonic	inharmonic
Frequencies:	stable	fluctuating
Pitch:	clear and salient	ambiguity of pitch(es)
Typical:	sensation of a single/dominant pitch	several pitches/no clear pitch

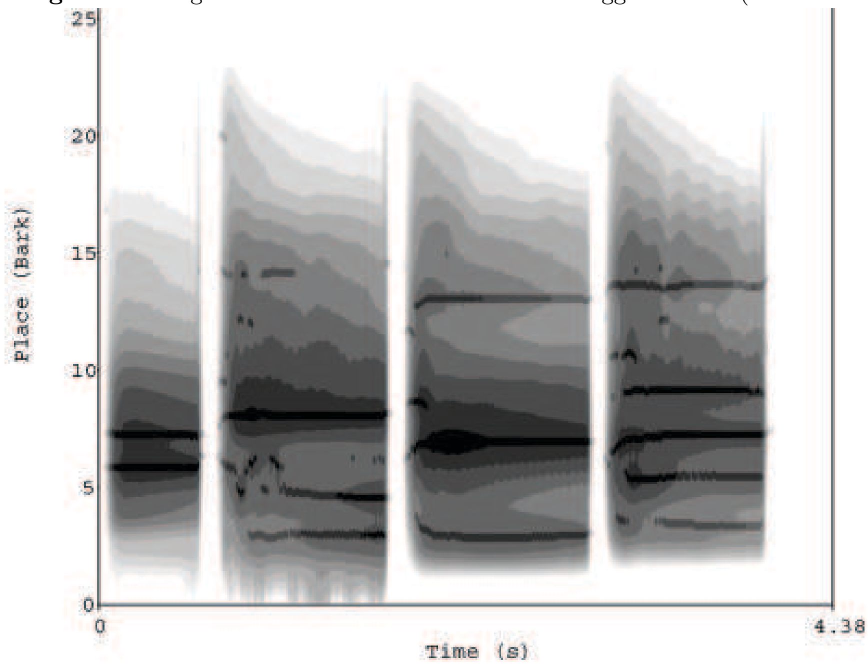
Most sounds produced from musical instruments can be ordered along these dimensions. Sounds from aerophones and chordophones thereby fall on the left side, sounds from membranophones and, in particular, idiophones predominantly on the right. Of course, the plucking of a string in a chordophone such as a harpsichord or guitar also results in a transitory and often quite inharmonic onset of a sound (Keiler et al. 2003, Bader 2005) before a quasi-stationary state in the vibrating system as well as in the sound radiated from such systems is reached. In idiophones (xylophones, metallophones; e.g., gongs and gong chimes, bells) many of which are set to vibration by an impulse there is no quasi-stationary regime in the vibrating system since the transitory onset is immediately followed by an often rapid decay of modes of vibration due to friction forces in the material set to motions as is obvious in many African and Asian xylophone types yet also in some metallophones (Schneider 1997b). Ambiguity of pitch perception increases with increasing inharmonicity of the sounds produced by idiophones such as found in the Javanese and Balinese *gamelan* (Schneider 1997b, 2000a/b, 2001). Perception is complicated by the fact that many sounds radiated from xylophones and metallophones (e.g., gong chimes of the Javanese *bonang* or Balinese *trompong* type) are quite short in duration, which means that pitch perception of complex inharmonic sounds must be achieved within a time span of, in many cases, 100-250 ms from onset.

In what follows, sounds recorded from certain idiophones will be analyzed by means of different software tools which in turn represent different models of peripheral auditory signal processing. In particular, an algorithm developed by Hermes (1988) based on the subharmonic matching process as proposed by Terhardt (1979, 1998) as well as an algorithm close to the concept of the harmonic sieve (Cohen et al. 1995) will be employed. Both models consider spectral properties of sounds and hence operate in the frequency domain. In contrast, the auditory model of Meddis and Hewitt (1991a/b) developed further by Meddis and O'Mard (1997, 2003) is based in the time domain. In several respects, it is similar to the Auditory Image Model (AIM) developed by Patterson et al. (1995) as well as to some other temporal approaches (de Cheveigné 2005). In the present study, the Auditory Modelling System (AMS, Meddis & O'Mard 2003) is used.

The analyses for the subharmonic estimation of pitch as well as for the estimations based on the harmonic sieve as implemented in the spatial pitch network model (SPINET; Cohen et al. 1995) have been performed by using routines included in the Praat environment (Version 5.0.38; Boersma & Weenink 2008). Though this software was designed for experimental phonetics, it can handle a broad range of musical sounds. As a first step of analysis, our sound examples have been processed in the frequency domain by means of a filter bank which simulates the excitation pattern on the BM, and which yields a cochleagram (scaled in Bark) as output. This type of analysis is useful for finding strong spectral components, which can be regarded as pitch candidates. If a cochleagram contains several such strong components (marked by dark lines in graphics based on greyscales), it is very likely that these will be perceived as separate spectral pitches, or that they interact in forming virtual pitches (cf. Terhardt 1979, 1998, Schneider 1997b, 2000a/b).

Let us begin with the cochleagram obtained from the sounds of four of the bells (nos. 1-4) of the Du Mery carillon of Brugge (Fig. 5). For the analysis, all sounds recorded close to the bells located on top of the belfry of Brugge[1] have been cut to segments of 0.6-1 s duration from the onset, and all selections have been normalized at -3 dB level. The cochleagram of the four sound segments represent a bell scale comprising the musical notes of *g*, *a*, *bb*, *c'*, whereby the prime (that is, the second partial of a typical minor-third bell; Schneider & Leman 2002) is taken as the decisive spectral component with regard to tuning and pitch.

Fig. 5. Cochleagram of four bell sounds from the Brugge Carillon (Bells no. 1 - 4)



In Tab. 2 the relevant modal frequencies determined by spectrographic analysis (amplitude levels omitted) for the strong partials of the four bell sounds in question are shown up to the double octave (a theoretical frequency ratio of 8:1 to the hum).

In Tab. 2, components a/b denote degenerate pairs of eigenmodes. The strong spectral components listed here as well as many more present in each sound radiated from one of the bells form the acoustical input to a bank of filters

[1] The equipment included two Neumann U 67, two TAB V72a preamps (fixed gain +32dB), one Telefunken V76 (variable gain 3-76dB), and a Panasonic SV 3800 DAT (48kHz/16 bit).

Table 2. Eigenfrequencies and frequency ratios with respect to the hum of bells no. 1-4 from the Brugge carillon

		No. 1 (g)		**No. 2** (a)		**No. 3** (bb)		**No. 4** (c')	
Partial	Name	f_n (Hz)	f_1/f_n	f_n (Hz)	f_1/f_n	f_n (Hz)	f_1/f_n	f_n (Hz)	f_1/f_n
0						65.13	0.50		
1	Hum	97.43	1.00	111.10	1.00	121.43	1.00	129.01	1.00
2	Prime	195.84	2.01	218.50	1.97	243.80	2.01	262.36	2.03
3	Tierce	233.81	2.40	261.60	2.35	292.77	2.41	309.62	2.40
4	Quint	294.51	3.02	328.57	2.96	361.88	2.98	393.61	3.05
5	Nominal	391.04	4.01	438.40	3.95	492.84	4.06	521.39	4.04
6a	10th	488.39	5.01	550.08	4.95	607.46	5.00	642.82	4.98
6b		492.86	5.05	569.38	5.12	627.23	5.16	666.34	5.16
7	11th	514.99	5.28	577.04	5.19	642.07	5.28	688.21	5.33
8		525.22	5.39	630.31	5.67			696.39	5.40
9a		573.35	5.88						
9b		576.26	5.91						
10	12th	589.41	6.05	661.34	5.95	741.58	6.01	763.68	5.92
								729.01	6.13
11	13th	627.59	6.44	755.07	6.32			847.23	6.57
12		683.73	7.01	701.85	6.80			923.12	7.15
13		712.67	7.31						
14		746.30	7.66					988.71	7.66
15a	Double	818.46	8.40	918.04	8.26	1027.05	8.46	1098.69	8.51
15b	Octave	824.06	8.46						

which produces the cochleagram as output. The excitation patterns (1-4) reflect that there are several strong components in each sound which could function as spectral pitches, and which might also contribute to the formation of virtual pitches.

The estimation of a single (low and/or dominant) pitch per bell sound with the Praat algorithm based on the SPINET model yields no interpretable result. The respective graph is empty for most of the duration of the four sound segments, and erratic for the remainder. This result may be surprising, or even disappointing because all four bell sounds contain a considerable number of quasi-harmonic components besides the inharmonic ones. One has to remember though that the spatial pitch network constitutes a weighted harmonic sieve, which yields a pitch estimate best if the spectral components of a stimulus correspond to a single harmonic series. In this respect, it closely resembles other spectral pattern matching (or harmonic template) models which can deal with harmonic spectra, and fail to assign a low (virtual) pitch to complex tones as inharmonicity increases to a point where the spectral components no longer fit well to one template or sieve. With increasing spectral inharmonicity, the error term calculated from the frequency deviations of the components relative to the center frequencies of a template or harmonic sieve surpasses a certain limit.

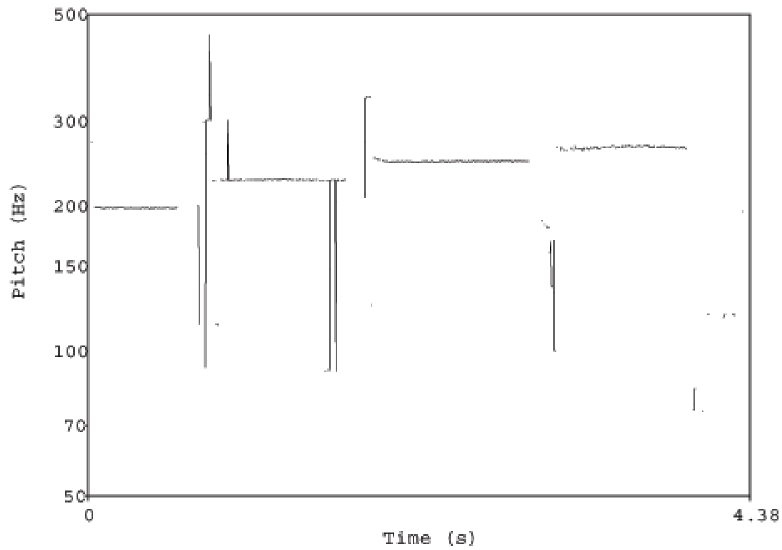
This implies that the probability for a certain spectral component to match a template or sieve decreases with increasing detuning of harmonic partials towards inharmonicity. For truly inharmonic sounds such as produced by *gamelan* instruments, the harmonic sieve model seems inappropriate for analysis because the distribution of spectral components along the frequency axis in such sounds usually is too complex, and too irregular to be matched to a harmonic series of frequencies (see below).

Applying the subharmonic matching model as devised by Hermes (1988) to the same four bell sound segments yields a much clearer result. The algorithm indeed produces one low pitch per sound (plus some artefacts, see Fig. 6), which in this case apparently represents the prime (2nd partial) of each sound.

The result of this analysis conforms to behavioral data which consistently demonstrate that many (if not most) subjects judge the main pitch of a bell to be equal to the so-called strike tone or strike note (German: *Schlagton*; Dutch: *Slagtoon*). The strike note is a virtual pitch typically located at or close to the frequency of either the second or the first partial of the bell spectrum (Terhardt & Seewann 1984). Since these two partials are an octave apart, yet equivalent as to their chroma, the results provided by the subharmonic matching model seem reasonable.

In a second trial, more complex sounds recorded in Bali by Rolf Bader from a *gendér wayang* (see Schneider 2001b) were fed into the SPINET as well as into the subharmonic matching model, respectively. Before that, the basic cochleagram analysis was carried out. The *gendér* used in this analysis is a metallophon that consists of ten bronze plates. A bamboo resonator is attached to each plate and is usually tuned so as to match a low spectral component of the plate. The gendér instruments are played in pairs labelled *pengumbang* and *pengisep* with

Fig. 6. Pitch of four bell sounds from the Brugge carillon, bell no. 1 - 4); Subharmonic matching



regard to their tuning. The lowest mode of each plate on the *pengumbang* is tuned 20-50 cents low relative to the respective plate and mode frequency on the *pengisep*. Since the two instruments are played so as to produce many intervals in parallel, this results in complex inharmonic sounds, which undergo AM permanently (for details, see Schneider 1997b, 2000a, 2001b). For the present analysis, the sounds recorded from plates no. 3, 4, 5 of the *pengumbang* played by a local musician in Bali have been used, which respresent part of a scale. The segments of these sounds subjected to analysis last about three seconds each. All sounds have been normalized at -6dB.

The cochleagram reveals that there are different zones of BM excitation, and that each sound segment offers several strong components which function as pitch candidates. In fact, at least two pitches can be identified by listeners in such sounds, namely a low pitch of a sinusoidal quality and a high pitch of metallic timbre. Depending on the strength of playing (Balinese music culture distinguishes between soft and strong styles of playing), the number of modes excited in each plate of course differs considerably, and so does the number of spectral and virtual pitches which are elicited. An analysis of the three sound segments with the SPINET model fails altogether. The subharmonic matching model yields the graph in Fig. 8 as output; frequency in this plot ranges from 50 Hz to 500 Hz (ordinate gives log frequency).

Fig. 7. Cochleagram of three sound segments recorded from a *gendér* (Bali)

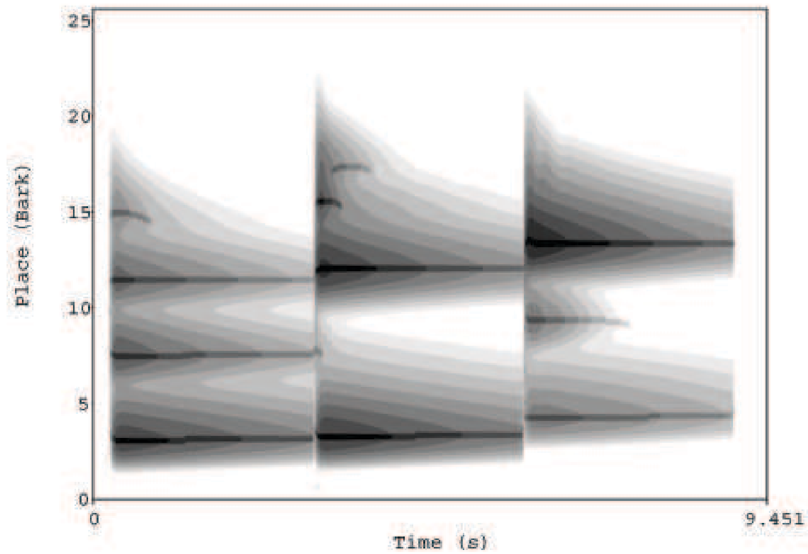
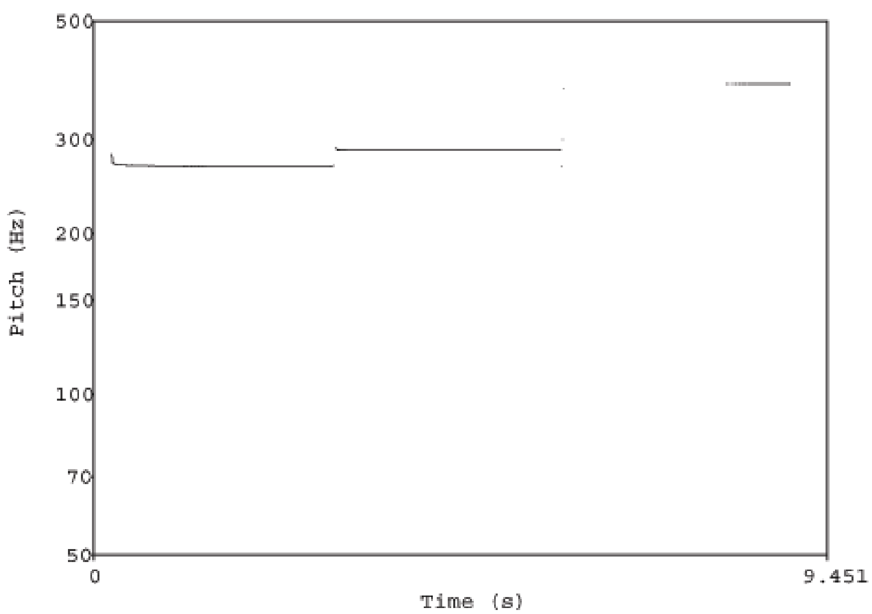


Fig. 8. Pitch estimates, 3 *gendér* sounds, subharmonic matching model



From Fig. 8 it is evident that the algorithm produces a single pitch estimate for each sound segment. In this case, these estimates do *not* correspond to low spectral components of the three sounds, which would have been found at 212.44 Hz, 247.7 Hz, and 286.6 Hz, respectively. Instead, the model yields three frequencies which seem to indicate virtual pitches. For the third sound segment, this frequency appears only towards the end of decay when spectral complexity is already reduced compared to the onset.

To check the performance of a model based in the time domain, we used the sound of a *bonang* kettle gong from West Java (Sunda; see Schneider 1997b), the sound from bell no. 1 of the Brugge carillon (Schneider & Leman 2002) as well as one of the *gendér* sounds as acoustic input. Processing of sound stimuli in the AMS (Meddis & O'Mard 1997, 2003) functionally includes the following basic steps: 1) peripheral excitation and BM filtering, 2) inner hair cell (IHC) model, half-wave rectifier, low-pass filtering, 3) extraction of periodicities within each channel by means of ACF, 4) aggregation of periodicities across channels with a summary ACF (SACF), which allows to find the pitch estimate of sounds. Since the AMS tries to emulate the auditory pathway as close as possible, it includes modules such as an outer/middle ear filter, conversion to stapes velocity, IHC/auditory nerve (AN) synapse, receptor potential, refractory period within a nerve fiber, etc. The model has been refined and expanded over the years to include a nonlinear BM (Lopez-Poveda & Meddis 2001) as well as other new features (see Meddis 2006). It is of importance to notice that the AMS model (after BM filtering and IHC transduction) operates on patterns of AN spike probabilities. Hence, the within-channel ACF as well as the SACF is not obtained from the waveshape of the sound yet from the neural activity patterns consequent to peripheral auditory stimulation.

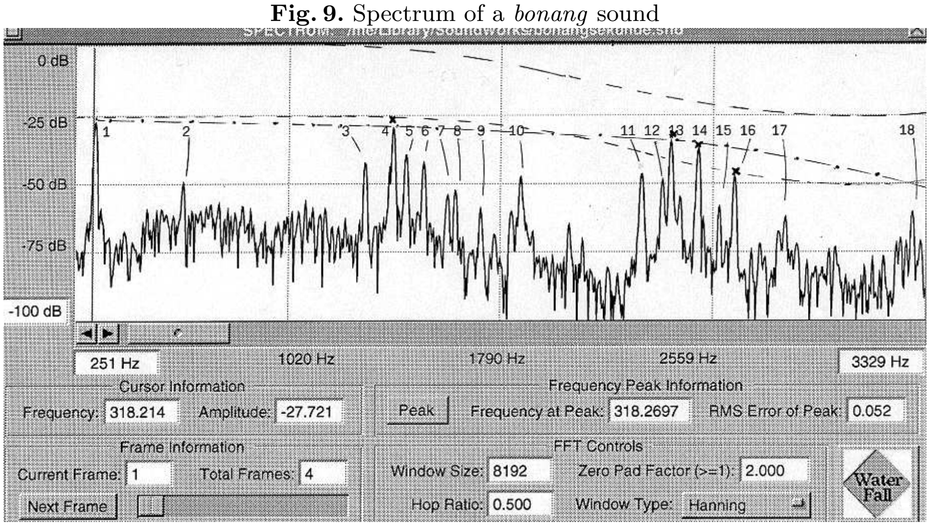
The main components of the spectrum of the *bonang* gong are listed in Tab. 3. The measurements were obtained with the 'Frequency at peak'-option (parabolic interpolation) of Spectro 3.01 (G. Scavone, P. Cook) running on a NeXT. The amplitudes are calculated relative to 0 dBfs; in an undistorted signal, all amplitudes must be ≤ 0 dBfs.

The spectrum of the *bonang* sound is shown in Fig. 9. The spectal envelope is indicated as a dash-dotted line, and contours of equal loudness as dashed lines. Peaks with amplitudes which can be expected to evoke a loudness sensation equal to, or greater than that of the lowest strong component at 318.3 Hz are marked with an 'x'. There are four relevant components (nos. 4, 13, 14, 16) in addition to the lowest one, which can be considered as candidates possibly giving rise to spectral pitches and/or contributing to a virtual pitch. This feature accounts for most of the ambiguity this sound brings about in regard to pitch perception. Another factor is that groups of narrowly spaced spectral components fall into the same critical band (CB). The respective CBs have been indicated in Tab. 3 according to data given in Zwicker & Fastl (1999, 159; Tab. 6.1). Groups of inharmonic spectral components falling into the same CB have two perceptual effects: first, they cause AM and roughness sensation; second, they further increase the pitch ambiguity of this peculiar sound. One has to remember that the

Table 3. Spectrum of *bonang* gong (Sunda); main spectral components (no. 1-16)

No.	f (Hz)	Rel. Ampl. (dB)	Bark (z)	Pitch Candidate
1	318.27	-27.7	3	X
2	638.30	-49.7	6	
3	1289.15	-42.5	10	
4	1401.48	-29.7	10	X
5	1448.85	-39.6	10	
6	1511.84	-42.8	11	
7	1599.07	-54.7	11	
8	1627.26	-52.5	11	
9	1719.67	-59.4	11	
10	1885.08	-47.7	12	
11	2304.28	-46.5	13	
12	2378.47	-49.2	14	
13	2408.19	-35.2	14	X
14	2507.20	-37.5	14	X
15	2583.97	-58.6	14	
16	2637.69	-47.7	14	X

Fig. 9. Spectrum of a *bonang* sound



bonang is a gong chime (usually comprising ten kettle gongs tuned to a scale; cf. Schneider 1997b) in the *gamelan* of Java which is often used to render the so-called nuclear theme, that is, a basic melodic pattern.

The *bonang* sound in question yields no interpretable result with the SPINET. Evidently, harmonic template or sieve models are unsuited to deal with such complex inharmonic spectra. Because of the spectral composition of this *bonang* sound, its time function cannot be expected to be sufficiently periodic to facilitate detection of a common period, T. The ACF calculated directly from the sound input indeed reveals no obvious period or regular fine structure.

Compared to a standard ACF, the AMS operates differently in that the signal is split into BM filter channels first[2], then undergoes transduction by means of an IHC and low pass filtering module before the within-channel periodicity, and finally the SACF is calculated. It could be that, nothwithstanding the CB problem obvious from the data in Tab. 3, such a stepwise processing might improve the result of each within-channel ACF, and thereby also the final SACF. The image obtained from the output of all within-channel ACFs indeed shows that for components which can be resolved individually by the peripheral filter bank, a period is found corresponding to the frequency of the component by $T_n = 1/f_n$. The image is difficult to analyse, though, for channels which contain the ACF calcaluted from several inharmonic components falling into the same filter band. A peak-picking algorithm applied to the SACF data finds major peaks at lags corresponding to 158.94, 350.36, 105.49, 108.11, 705.88, 470.59, 128.69, 205.13, and 237.62 Hz, respectively. The frequencies, which are ordered according to peak height expressing $r_{xx'}$, do not correspond to spectral components (though 350.36 Hz is relatively close to the lowest modal frequency of the *bonang*). One could hypothesize that a peak around 9.4 ms ($\sim$106 Hz) is representing a near-periodicity, which can be established from subharmonic matching of several of the spectral components of the *bonang* (in particular, nos. 1-3; cf. Tab. 3). Since there are peaks in the SACF at 105.5 and 108.1 Hz, respectively, a quasi-period of $\sim$106 Hz, which (as a common denominator) fits to a number of spectral components, seems feasable. Fig. 10 shows the SACF of the *bonang* sound for 100 ms.

For the first of the carillon bells, the same analysis in the time domain was carried out. Due to the much less inharmonic, in parts quasi-harmonic spectral structure of the bell sound (see Tab. 2), the resulting SACF contains a number of small yet distinct peaks (Fig. 11).

A peak-picking algorithm applied to the SACF data detects the time point for each of the smaller peaks. Those listed in the following table have been ordered with respect to their relative height (expressing the degree of correlation $r_{xx'}$)

[2] For the present analysis, a gammatone filter bank (with CF from 100 to 4000 Hz) has been used because of the moderate signal level of the sounds (carillon bell, *bonang*, *gendér*). Otherwise, the nonlinear BM model (Lopez-Poveda & Meddis 2001) available in the AMS would have been employed.

Fig. 10. SACF (100 ms), *bonang* sound

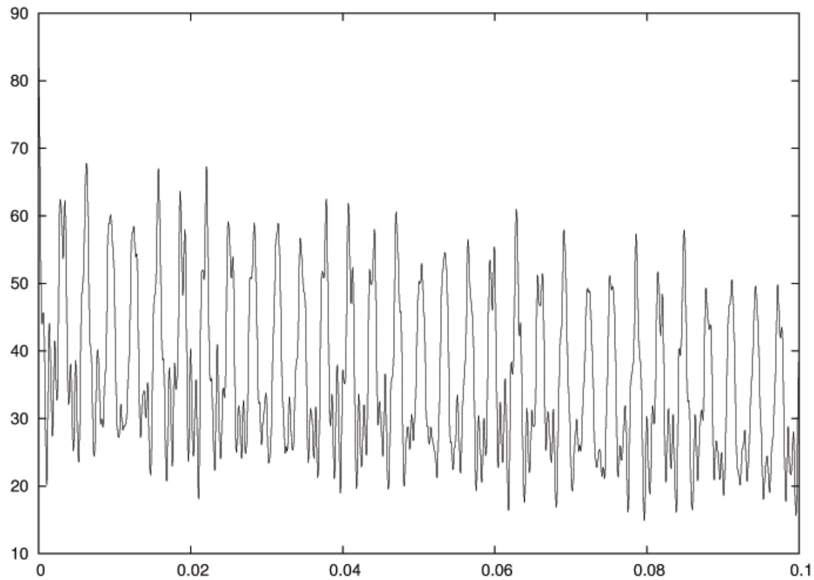
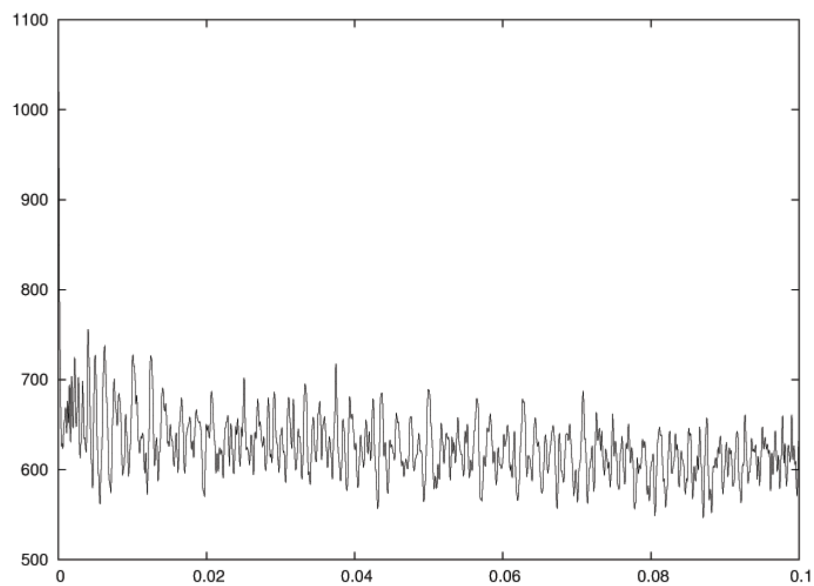


Fig. 11. SACF (100 ms), sound from carillon bell no. 1, Brugge



and the time lag (τ). In addition, the frequency corresponding to the lag due to $f = 1/\tau$ for each peak is given.

Table 4. Peaks found in the SACF of the bell sound ordered according to height

No.	Rel. Height	τ (ms)	f (Hz)
1	755.88	4.02	248.70
2	737.85	6.25	160.00
3	727.17	10.10	99.38
4	727.05	4.98	200.80
5	726.38	12.50	80.13
6	724.86	2.20	452.83
7	717.55	37.50	26.68
8	714.90	10.30	97.56

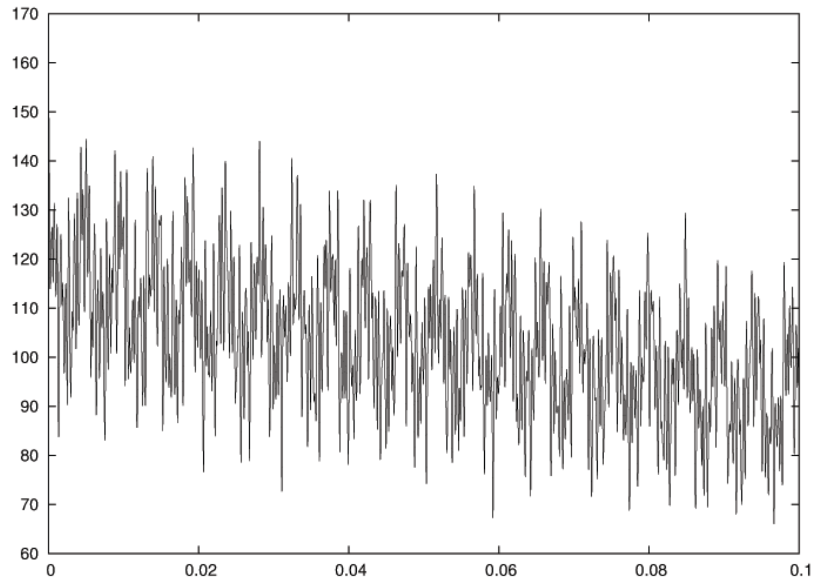
Some of the data can be interpreted (at least hypothetically) thus: peaks 3 and 8 are very close to the frequency of the hum note of this bell, that is, they might indicate the period corresponding to the lowest spectral component of the bell. Peak no. 4 comes close to the period and frequency of the prime. The strongest peak (no. 1 at 4.02 ms) might indicate some period as well since there are peaks also at near-multiples of its lag, that is, at 8.19, 12.5, 16.6, 20.7, 25.1, and 29.2 ms.

If the peaks are ordered according to the time lags $\tau < 12.5$ ms (which means $f > 80$ Hz), leaving aside the values for the peak height, the following frequencies are found: 80.13, 84.66, 87.91, 89.55, 94.5, 97.56, 99.38, 109.84, 122.14, 132.6, 160, 200.84, 248.7, 280.7, 303.8, 369.23, 452.83, 558.14, 666.66, 800, 1021.28,and 1920 Hz. Of these frequency values (which fit an exponential function $e^{0.33f}$ fairly well), those which are in the range of ca. 85-100 Hz seem to indicate that there is a periodicity of the partly harmonic, partly inharmonic signal which roughly corresponds to the lowest spectral component.

Finally, one of the *gendér* sounds (plate/scale step no. 3) was subjected to the same type of analysis. The SACF (Fig. 12) in this case is particularly complex due to the dense inharmonic spectrum, which contains strong peaks from 212 Hz up to 6.6 kHz, and moreover exhibits two maxima in the spectral envelope at ca. 2.6 and 6.6 kHz, respectively (cf. Schneider 2001b, Fig. 4).

A peak-picking algorithm applied to the SACF data finds the strongest peak at a lag of 5.02 ms which corresponds to 199.17 Hz. This is not too far away from the base mode frequency of the plate (~212.5 Hz), however, a multitude of relatively strong peaks in the SACF makes it unlikely that a common period could be extracted (a cluster analysis applied to the SACF data did not help much to clear the picture). Due to their spectral and temporal structure, *gendér* sounds are ambiguous in pitch. Typically, one can assign (for example, by at-

Fig. 12. SACF (100 ms), sound of *gendér* plate no. 3



tuning a tone generator or simply by singing a syllable such as "la") a main (low) pitch to each *gendér* sound; in experiments, subjects try to match the frequency of a sine tone (or the fundamental of the syllables they sing) to that spectral component of the *gendér* sound which represents the base mode of the transversal bending waves of the vibrating plate. Detecting this low component is possible if it can be resolved by peripheral filtering (a condition which applies to our example), and probably processed individually after mechano-electrical transduction (as the output from within-channel periodicity extraction in the AMS module indicates).

Perceiving a low pitch, though, in most *gendér* (and, similarly, *bonang* and *trompong*) sounds goes along with perceiving from one to three additional components in higher frequency regions. If such components result from a single strong spectral peak unhampered by neighbours, they can mostly be interpreted as 'side pitches' complementing a 'main' (low) pitch. Because of the inharmonic composition of the spectrum, main pitch and side pitches do not fuse, that is, they are perceived as being more or less unrelated. In this respect, the perceptual situation is much different from a complex harmonic sound (see above, section 3). If groups of inharmonic spectral components fall into the same CB in frequency regions much higher than the base frequency, they may not give rise to a clear side pitch yet, rather, to a sensation of a cluster-like band of frequencies. This sensation often interferes with the percept of the low pitch, which thereby appears less salient. Together, low pitch plus side pitches, or low pitch plus spectral disturbances (plus some FM as well as a considerable degree of AM) make up a situation where perceptual ambiguity is almost inevitable.

One of the reasons why the *gendér* sounds are so awkward with regard to pitch perception, seems to be the geometry of the plates, which have a trapezoid cross section. Wave dispersion within this geometry allows for more complex patterns of reflections at boundaries, which in turn result in rich inharmonic sounds in particular during the transient portion (Bader 2004 and personal communication). Since the plates are hit with mallets, and the number of notes played per time unit in the *gong kebyar* style often is high, listeners are confronted with fast sequences of transient inharmonic sounds, to the effect that pitch perception for each short sound segment as well as for sonorities formed from several sounds produced simultaneously will be ambiguous.

5 Discussion

Comparing the results from the two models operating in the frequency domain to the AMS operating in the time domain, it seems reasonable to state that all can handle periodic signals where they detect f_0 and/or spectral components of the actual sound input. In regard to inharmonic sounds, both the subharmonic matching model and the AMS based on the ACF approach can analyze signals which are not strictly periodic, and which have a partly harmonic, partly inharmonic spectrum like, for example, carillon bells such as found at Brugge. To be sure, these bells are excellent specimen which have been carefully tuned both with regard to their spectral composition as well as constituting a musical scale when played one after another.

The results obtained from computer-based ear models are in line with many observations and experiments which demonstrate that with increasing spectral inharmonicity and the periodicity of the signal decreasing correspondingly, the pitch or, rather, pitches assigned by listeners to such sounds as produced by bells, gong chimes etc., typically become ambiguous (Terhardt, Stoll & Seewann 1982b, Terhardt & Seewann 1984, Schneider 1997b, 2000a/b, 2001a/b).

In one experiment, in which we employed samples from bell no. 2 of the Brugge carillon for the reproduction of a chorale (*Ich bin's, ich sollte büßen…*, J. S. Bach; four voice transcript for piano or organ), apperception even for skilled listeners was difficult. Even though the fundamental frequencies of all notes realized with the bell sounds were identical with the fundamental frequencies as defined by the notation of the given piece, the spectral inharmonicity of the bell led to severe problems in analyzing the musical structure, which consists of chords played one after another. With each chord comprising four bells sounds (one for each voice), the resulting spectral inharmonicity is considerable, and evidently hampers detection of the pitches, which is necessary for listeners to be able to follow the motion of the four voices, on the one hand, and to understand the tonal functions of the simultaneous chords, on the other. In regard of variables related to perception (such as consonance, roughness, etc.), the bell version of the chorale differs significantly from one played with a (synthesized) pipe organ (Schneider 2001a).

It is quite obvious that perceptual and cognitive processing of complex inharmonic sounds, and music played with such sounds is more difficult and demanding than is perception of harmonic complex tones as well as listening to music based on such. As has been observed in experiments with evoked potentials, several parameters (such as the latencies for P1 and N1) change significantly when inharmonic instead of harmonic sounds are used as stimuli (cf. Sinex 2005, 380-81).

The ambiguity of pitch experienced in complex inharmonic sounds such as radiated from bells and gongs can be attributed to both the spectral composition and the temporal structure of such stimuli. Though spectral and temporal structure interacts in may ways, they should be examined separately. For example, many sounds from carillons and *gamelan* instruments, due to the inharmonicity and density of spectral components (for examples, see Schneider 1997b, 2000a/b, 2001a/b, Schneider & Leman 2002) mean an increased workload of spectral processing by the auditory system. Even though in many stimuli some of the strong components can be resolved (according to CB filter bands), and may thus serve as spectral cues for pitch perception, they often form arbitrary frequency ratios, and hence give rise to separate spectral pitches. Also, one finds inharmonic sounds in which some of the strong spectral components are closely spaced in frequency, and thereby interact both as spectral pitch candidates as well as producing AM.

Given such a spectral structure, the corresponding time function cannot be periodic, and the auditory system in such cases fails to extract a basic periodicity corresponding to f_0 as temporal information relevant for pitch perception. If several inharmonic sounds are played simultaneously (as is the case in *gamelan* music and also in music played on a carillon), ambiguity increases as a function of the resulting spectral inharmonicity and density, which hampers identification of musically relevant objects (e.g., motives, themes, phrases) within each voice as well as separation of voices in chords and sonorities.

It is not possible, at this point, to discuss the physiological relevance and validity of various auditory models as well as objections which have been raised against approaches centered either in the frequency or in the time domain (see de Boer 1976, Lyon & Shamma 1996, Terhardt 1998, Lopez-Poveda 2005, Plack & Oxenham 2005, de Cheveigné 2005). In regard to spectral models, the neglect of information available from the temporal envelope and its fine structure has often been critized. On the other hand, objections against a purely temporal approach to pitch and interval perception repeatedly have pointed to the weak neuroanatomical and neurophysiological foundations of, in particular ACF-based models. Empirical evidence for the validity of the ACF-approach has been sought in interspike-interval codes recorded from the AN where data allow to regard so-called all-order interspike interval histograms (ISIH), in particular, if pooled from a sample of AN fibers, as being “autocorrelation-like neural representations of the stimulus" (Cariani & Delgutte 1996a, 1712). Basically the same approach has been used more recently to show that the ISIH obtained in experiments on AN fibers of the cat for consonant musical intervals, namely the perfect fifth (ratio of the two fundamental frequencies 660 Hz : 440 Hz = 3:2) and the perfect fourth (4:3) entails perfect periodicity (Tramo et al. 2001). For the more dissonant interval of the tritone (45:32), the ISIH is less clear in regard to periodicity, and for the minor second (469 Hz : 440 Hz $\approx$ 16:15), periodicity is more difficult to extract from the ISIH because the first strong peak occurs only after a lag $\tau \sim$35 ms.

The fact that a periodic sound stimulus such as a vowel can trigger a periodic spike response synchronized to the phase of a tone has been known for decades. Also, it was shown that the spectrum of the input signal can be reconstructed from the neural discharge pattern in AN fibers. There have been a number of observations with regard to nuclei and mechanisms along the auditory pathway, where neural periodicity coding and analysis might be performed (cf. Keidel 1992, Ehret 1997, de Ribaupierre 1997, Schneider 1997b, 84-95, 104-109, 135-142; 2000a). Further, it has been argued that stimulus periodicity is mapped in the auditory cortex (AI; see Schulze et al. 2002). Recently, observations as well as hypothetical interpretations have been condensed into a temporal model of pitch and timbre perception as well as the neural basis of harmonicity detection (Langner 2007).

The concept of periodicity is so fundamental to natural processes such as vibration and sound that it would be a surprise if the sense of hearing in mammals would not be capable to detect such periodicities, and to use them as temporal cues for pitch and interval perception (see also Keidel 1989, Yost 2004). From the empirical evidence available, it can be inferred that most if not all of the neural processing necessary to determine the low pitch of signals as well as to perceive basic musical intervals, is achieved in subcortical networks (mainly of the brainstem and thalamus).

The inharmonic variety of sounds, which no doubt is also a natural phenomenon as evident from many environmental sounds, can be viewed as a deviation from the basic, periodic as well as harmonic situation. In experiments, deviations of periodicity and harmonicity of sounds can be realized sytematically and continuously in order to study the perceptual effects of, for example, mistuning of individual partials, or of regular spaced yet inharmonic partials etc. (cf. Roberts 2005). With respect to musical instruments and their sounds viewed from a transcultural perspective, one finds many degrees of inharmonicity as well as sounds with weak spectral and temporal coherence. Ambiguity of pitch and timbre, which in such sounds is a natural consequence to their temporal and/or spectral organization, in many musical cultures however is a feature that apparently is wanted. For examples, one might point to the *gong kebyar* style of Balinese *gamelan* music with a multitude of inharmonic sonorities as well as to plenty of AM resulting from the difference in tuning between *pengum-*

bang and *pengisep* metallophones mentioned above[3]. With carillon music, the degree of overall inharmonicity perhaps is smaller compared to a *gamelan gong kebyar* performance, however, the ambiguity of pitches produced by each single bell as well as by sequences of bell sounds in principle is the same as is with metallophones from Java or Bali. With carillons, a particularly difficult task is to derive the tonal meaning of chords or even rapid chord progressions played on such instruments[4]. Even a single major chord played on a carillon, due to the spectral composition of so-called minor-third bells, which are still predominant in European carillons, often appears indifferent with respect to tonality (that is, somewhere in between major and minor). This situation in which the acoustic sound structure is not identical with the musical structure of many of the pieces played on such carillons causes kind of 'friction' as well as perceptual ambiguity, which, to be sure, poses problems to analytical listening and music cognition. However, such a sound structure can also be appreciated as an enrichment of our perceptual and cognitive experience.

Whereas musically trained listeners in general will succeed in perceiving and categorizing major or minor chords played on conventional chordophones, for example a piano, they mostly have difficulties performing such tasks when the stimuli come from a (real or sampled) carillon. Judging the size of intervals or the structure of chords played simultaneously for such stimuli is demanding due to fact that bell sounds rarely give rise to simple and unambiguous pitch percepts. Rather, one has to deal with two or more (spectral and/or virtual) pitches per bell sound (Terhardt & Seewann 1984, Terhardt 1998). Listening analytically to carillon music therefore always includes the task of reducing the complex acoustical input (which implies high dimensionality in regard to perception and cognition) to a simpler, musically relevant representation reduced in dimensionality. This task is complicated if pieces are played in fast tempo, and in a virtuos performing style, which makes it difficult if not impossible to perform analytical listening in quasi-real time. Of course, training will improve the performance of listeners who try to analyze musical structures from pieces played on a carillon. Also, there are Gestalt effects in that well-known melodies will be identified acccording to their basic melodic and rhythmic patterns notwithstanding the unusual sound quality of bells.[5]

---

[3] An illustrative and easily available example is found on the CD *Music from the morning of the world*, rec. by David Lewiston in Bali (Elektra/Asylum/Nonesuch Records 1988), no. 2 (*gamelan gong kebyar: baris; gambang betjak*). Also, *The Music of Bali*, rec. by David and Kay Parsons (celestial harmonies 1997), offers good examples (e.g., *Kebyar trompong, Ujan Mas*).

[4] Good examples are found on the CD *Torenmuziek Dordrecht. Beiaardmuziek van Staf Nees. Carillon-recital Grote Kerk Dordrecht.* Henry Groen & Boudewijn Zwart (TMD 2001).

[5] Very instructive examples from *Lili Marleen*, and Turlough O'Carolan's *Lord Inchiquin*, to familiar Christmas charols can be found on the CD *Folk songs and popular music on European carillons* (Various artists, Eurocarillon 2000).

6 Conclusion

In the present study, algorithms which model parts of the auditory system have been tested on harmonic and inharmonic sounds in order to demonstrate that perceptual qualities such as salience or ambiguity of pitch, sensation of harmonicity and consonance as well as roughness and dissonance, are established on a relatively low level of auditory processing. Though we by no means underestimate the role of top-down processing (which employs natural or learned categories, schemas, memory etc.), we want to emphasize biological foundations of hearing, which govern perceptual processes up to the point of generating musical meaning. In music, meaning basically rests in specific sound structures as well as in relations sounds form among each other. Listening to music, from an ecological point of view, is not just a matter of cognitive activity leading to some mental representations of abstract musical structures (see also Clarke 2005, ch. 1); rather, one has to see that much of the neural processing in particular of musical sound and speech which are carrying information is achieved subcortically, and that musicians even seem to dispose of enhanced subcortical processing capabilities (cf. Musacchia et al. 2007). Of course, listening to music, in particular in an attentive mode, requires cognitive analysis and awareness. However, perception of inharmonic sounds, and of music composed of such sounds, clearly demonstrates that even for musically trained subjects analytical skills cannot easily overcome the constraints imposed by acoustic and psychoacoustic parameters.

Acknowledgements

We would like to thank Aimé Lombaert (Brugge) and Marc Leman (Ghent) for support in recording the bells at Brugge used in this analysis, and Rolf Bader (Hamburg) for providing some of his field recordings from North Bali.

References

1. Bachmann, Wolfgang 1992. Signalanalyse. Grundlagen und mathematische Verfahren. Braunschweig: Vieweg.
2. Bader, Rolf 2004. Additional modes in transients of a Balinese gender dasa plate. Journal of the Acoust. Soc. of America 116, 2621 (Abstract).
3. Bader, Rolf 2005. Computational Mechanics of the classical guitar. Berlin: Springer.
4. Balaban, Mira, Kemal Ebcioglu, Otto Laske (eds.) 1992. Understanding Music with AI: Perspectives on Music cognition. Cambridge, MA, Menlo Park: The AAAI Pr./MIT Pr.
5. Boersma, Paul 1993. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise-ratio of sampled sound. Proc. of the Inst. of Phonetic Sciences [Univ. of Amsterdam] 17, 97-110.
6. Boersma, Paul, David Weenink 2008. Praat: doing phonetics by computer. Version 5.038.
7. Bregman, Albert 1990. Auditory Scene Analysis. The perceptual organization of sound. Cambridge, MA: MIT Pr.

8. Cariani, Peter, Bertrand Delgutte 1996. Neural Correlates of the pitch of complex tones. I: Pitch and pitch salience. Journal of Neurophysiology 76, 1698-1716.
9. Cariani, Peter, Bertrand Delgutte 1996. Neural Correlates of the pitch of complex tones. II: Pitch shift, pitch ambiguity, phase invariance, pitch circularity, and the dominance region for pitch. Journal of Neurophysiology 76, 1717-1734.
10. Clarke, Eric 2005. Ways of Listening. A ecological approach to the perception of musical meaning. London: Oxford U.Pr.
11. Cohen, Michael, Stephen Grossberg, Lonce Wyse 1995. A spectral network model of pitch perception. Journal of the Acoust. Soc. of America 98, 862-879.
12. de Boer, E. 1976. On the "Residue" and auditory pitch perception. In W.D. Keidel, W. D. Neff (eds.). Handbook of Sensory physiology, Vol. V,3, Berlin, New York: Springer, ch. 13.
13. de Cheveigné, Alain 2005. Pitch perception models. In Chr. Plack et al. (eds.). Pitch. Neural coding and perception. 169-233.
14. de Ribaupierre, Franois 1997. Acoustical information processing in the auditory thalamus and cerebral cortex. In G. Ehret, R. Romand (eds.). The Central Auditory System. New York, Oxford: Oxford U.Pr., ch. 5, 317-388.
15. Ehret, Günther 1997. The auditory midbrain, a "shunting yard" of acoustical information processing. In G. Ehret, R. Romand (eds.). The Central Auditory System. New York, Oxford: Oxford U.Pr., ch. 4, 259-316.
16. Handel, Stephen 1989. Listening. An Introduction to the perception of auditory events. Cambridge, MA: MIT Pr.
17. Hartmann, William 1998. Signals, Sound, and Sensation. New York: Springer.
18. Hermes, Dik 1988. Measurement of pitch by subharmonic matching. Journal of the Acoust. Soc. of America 83, 257-264.
19. Hesse, Horst-Peter 1972. Die Wahrnehmung von Tonhöhe und Klangfarbe als Problem der Hörtheorie. Köln: A. Volk.
20. Hesse, Horst-Peter 1982. The Judgment of musical intervals. In M. Clynes (ed.). Music, mind and brain. New York, London: Plenum Pr., 217-225.
21. Houtsma, Adriaan 1995. Pitch perception. In B. Moore (ed.). Hearing. 2nd ed. San Diego, London: Academic Pr., 267-295.
22. Keidel, Wolf-Dieter 1989. Biokybernetik des Menschen. Darmstadt: Wiss. Buchges.
23. Keidel, Wolf-Dieter 1992. Das Phänomen des Hörens. Ein interdisziplinärer Diskurs. Naturwissenschaften 79, 300-310, 347-357.
24. Keiler, Florian, Can Karadogan, Udo Zölzer, Albrecht Schneider 2003. Analysis of transient musical sounds by auto-regressive modeling. Proc. 6th Intern. Conf. on Digital Audio Effects DAFx-03, London: Queen Mary, Univ. of London, 301-304.
25. Krumhansl, Carol 1998. Cognitive Foundations of musical Pitch. Oxford, New York: Oxford U.Pr.
26. Langner, Gerald 2007. Die zeitliche Verarbeitung periodischer Signale im Hörsystem: Neuronale Repräsentation von Tonhöhe, Klang und Harmonizität. Zeitschrift für Audiologie 46, 8-21.
27. Leman, Marc 1995. Music and Schema theory. Cognitive foundations of systematic musicology. Berlin: Springer.
28. Lopez-Poveda, Enrique 2005. Spectral processing by the peripheral auditory system: facts and models. Intern. Rev. of Neurobiology 70, 7-48.
29. Lopez-Poveda, Enrique, Ray Meddis 2001. A human nonlinear cochlear filterbank. Journal of the Acoust. Soc. of America 110, 3107-3118.
30. Lyon, Richard, Shihab Shamma 1996. Auditory representations of timbre and pitch. In H. Hawkins, T. McMullen, A. Popper, R. Fay (eds.), Auditory Computaton. New York: Springer, ch. 6, 221-270.

31. McAdams, Stephen, Emmanuel Bigand (eds.) 1993. Thinking in Sound. The Cognitive psychology of human audition. Oxford: Clarendon Pr.
32. Meddis, Ray, Michael Hewitt 1991a. Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification. Journal of the Acoust. Soc. of America 89, 2866-2882.
33. Meddis, Ray, Michael Hewitt 1991b. Virtual pitch and phase sensitivity of a computer model of the auditory periphery. II: Phase sensitivity. Journal of the Acoust. Soc. of America 89, 2883-2894.
34. Meddis, Ray, Lowel O'Mard 1997. A unitary model of pitch perception. Journal of the Acoust. Soc. of America 102, 1811-1820.
35. Meddis, Ray, Lowel O'Mard 2003. AMS Tutorial (Version 2.3). Colchester: Univ. of Essex, Dept. of Psychol.
36. Meddis, Ray 2006. Auditory-nerve first-spike latency and auditory absolute threshold: a computer model. Journal of the Acoust. Soc. of America 119, 406-417.
37. Musacchia, Gabriella, Mikko Sams, Erika Skoe, Nina Kraus 2007. Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. Proc. Nat. Acad. Science Vol. 104, no. 40, 15894-15898.
38. Patterson, Roy, Mike Allerhand, Christian Giguère 1995. Time-domain modeling of peripheral auditory processing: a modular architecture and a software platform. Journal of the Acoust. Soc. of America 98, 1890-1894.
39. Plack, Christopher, Andrew Oxenham 2005. The Psychophysics of pitch. In Chr. Plack et al. (eds.). Pitch. Neural coding and perception. New York: Springer, 7-55.
40. Plack, Christopher, Andrew Oxenham, Richard Fay, Arthur Popper (eds.) 2005. Pitch. Neural Coding and Perception. New York: Springer.
41. Popper, Arthur, Richard Fay (eds.) 1992. The Mammalian Auditory Pathway: Neurophysiology. Berlin, New York: Springer.
42. Rameau, Jean-Philippe 1722. Traité de l'harmonie. Paris: Ballard.
43. Rameau, Jean-Philippe 1737. Génération harmonique ou traité de musique théorique et pratique. Paris: Prault fils.
44. Roberts, Brian 2005. Spectral pattern, grouping, and the pitches of complex tones and their components. Acta Acustica united with Acustica 91, 945-957.
45. Schneider, Albrecht 1997a. *Verschmelzung*, tonal fusion, and consonance: Carl Stumpf revisited. In Marc Leman (ed.). Music, Gestalt, and Computing. Studies in Cognitive and Systematic Musicology. Berlin, New York: Springer, 117-143.
46. Schneider, Albrecht 1997b. Tonhöhe - Skala - Klang. Akustische, tonometrische und psychoakustische Studien auf vergleichender Grundlage. Bonn: Orpheus-Verlag.
47. Schneider, Albrecht 2000a. Inharmonic Sounds: implications as to 'Pitch', 'Timbre' and 'Consonance'. Journal of New Music Research 29, 275-301.
48. Schneider, Albrecht 2000b. Virtual Pitch and musical instrument acoustics: the case of idiophones. In B. Enders, J. Stange-Elbe (eds.). Musik im virtuellen Raum. KlangArt-Kongreß 1997. Osnabrück: Rasch, 397-417.
49. Schneider, Albrecht 2001a. Complex inharmonic sounds, perceptual ambiguity, and musical imagery. In Rolf Inge Godøy, Harald Jørgensen (eds.). Musical Imagery. Lisse, Abingdon: Swets & Zeitlinger, 95-116.
50. Schneider, Albrecht 2001b. Sound, Pitch, and Scale: From "Tone measurements" to sonological analysis in ethnomusicology. Ethnomusicology 45, 489-519.
51. Schneider, Albrecht 2008. Foundations of Systematic Musicology: a study in history and theory. In A. Schneider (ed.). Systematic and Comparative Musicology: concepts, methods, findings. Frankfurt/M.: P. Lang, 11-61.

52. Schneider, Albrecht, Marc Leman 2002. Sonological and psychoacoustic Characteristics of carillon bells. In M. Leman (ed.). The quality of bells. Brugge Eurocarillon 2002 (=Proc. Of the 16th Meeting of the FWO Research Soc. on Foundations of Music research). Ghent: IPEM, Univ. of Ghent.
53. Schneider, Albrecht, Rolf Bader 2003. Akustische Grundlagen musikalischer Klänge. Mitteilungen der Math. Ges. in Hamburg Bd XXII, 27-44.
54. Schulze, Holger, Heinrich Neubauer, Frank Ohl, Andreas Hess, Henning Scheich 2002. Representation of stimulus periodicity in the auditory cortex: recent findings and new perspectives. Acta Acustica united with Acustica 88, 399-407.
55. Sinex, Donald 2005. Spectral processing and sound source determination. International Review of Neurobiology 70, 371-398.
56. Sloboda, John 1985. The musical Mind. The Cognitive Psychology of Music. Oxford: Clarendon Pr.
57. Stumpf, Carl 1890. Tonpsychologie, Bd 2. Leipzig: J. Barth.
58. Stumpf, Carl 1926. Die Sprachlaute. Berlin: J. Springer.
59. Terhardt, Ernst 1989. Calculating virtual pitch. Hearing Research 1, 155-182.
60. Terhardt, Ernst 1998. Akustische Kommunikation. Berlin, New York: Springer.
61. Terhardt, Ernst, Manfred Seewann 1982a. Pitch of complex signals according to virtual-pitch theory: tests, examples, and predictions. Journal of the Acoust. Soc. of America 71, 671-678.
62. Terhardt, Ernst, Manfred Seewann 1982b. Algorithm for extraction of pitch salience from complex tonal signals. Journal of the Acoust. Soc. of America 71, 679-688.
63. Terhardt, Ernst, Manfred Seewann 1984. Auditive und objektive Bestimmung der Schlagtonhöhe von historischen Kirchenglocken. Acustica 54, 129-144.
64. Tramo, Mark, Peter Cariani, Bertrand Delgutte, Louis Braida 2001. Neurobiological Foundations for the theory of harmony in Western tonal music. In Robert Zatorre and Isabelle Peretz (eds.). The Biological Foundations of music (=Annals of the N.Y. Acad. of Sciences, Vol. 930). New York: New York Acad. of Sciences, 92-116.
65. Wever, Ernest 1949. Theory of hearing. New York: Wiley.
66. Wiener, Norbert 1961. Cybernetics or control and communication in the animal and in the machine. 2nd ed. New York: MIT Pr.
67. Yost, William 2004. Determining an auditory scene. In M. Gazzaniga (ed.). The Cognitive Neurosciences. 3rd ed. Cambridge, MA, London: MIT PR., ch. 28, 385-396.
68. Zwicker, Eberhard, Hugo Fastl 1999. Psychophysics. Facts and Models. 2nd ed. Berlin, New York: Springer.