

Exploring the temporal mechanism involved in the pitch of unresolved harmonics

Christian Kaernbach^{a)} and Christian Bering

Institut für Allgemeine Psychologie, Universität Leipzig, Seeburgstr. 14-20, 04103 Leipzig, Germany

(Received 3 September 1999; accepted for publication 2 May 2001)

This paper continues a line of research initiated by Kaernbach and Demany [J. Acoust. Soc. Am. **104**, 2298–2306 (1998)], who employed filtered click sequences to explore the temporal mechanism involved in the pitch of unresolved harmonics. In a first experiment, the just noticeable difference (jnd) for the fundamental frequency (F_0) of high-pass filtered and low-pass masked click trains was measured, with F_0 (100 to 250 Hz) and the cut frequency (0.5 to 6 kHz) being varied orthogonally. The data confirm the result of Houtsma and Smurzynski [J. Acoust. Soc. Am. **87**, 304–310 (1990)] that a pitch mechanism working on the temporal structure of the signal is responsible for analyzing frequencies higher than ten times the fundamental. Using high-pass filtered click trains, however, the jnd for the temporal analysis is at 1.2% as compared to 2%–3% found in studies using band-pass filtered stimuli. Two further experiments provide evidence that the pitch of this stimulus can convey musical information. A fourth experiment replicates the finding of Kaernbach and Demany on first- and second-order regularities with a cut frequency of 2 kHz and extends the paradigm to binaural aperiodic click sequences. The result suggests that listeners can detect first-order temporal regularities in monaural click streams as well as in binaurally fused click streams. © 2001 Acoustical Society of America. [DOI: 10.1121/1.1381535]

PACS numbers: 43.66.Ba, 43.66.Hg, 43.66.Mk [RVS]

I. INTRODUCTION

A temporal mechanism involved in the perception of pitch has been suspected since the days of Ohm and Seebeck. In 1940, Schouten had suggested the existence of a mechanism which derives the pitch of unresolved harmonics through an analysis of the periodicity of the waveform (Schouten, 1940, 1970). In discrimination tasks with band-pass filtered pulse trains, Hoekstra (1979) measured a pitch jnd of below 0.1% for stimuli which contained harmonics below the 10th, while obtaining a jnd of 2% for harmonic orders above 20. In between, he found a transition. He argued that these levels reflect two different mechanisms of pitch analysis, one working for stimuli with spectrally resolvable components, and one working entirely on the temporal structure of the stimulus. His results were replicated by Houtsma and Smurzynski (1990) who showed that pitch discrimination for complex tones comprising high harmonics added in sine phase functions on two levels of performance. While the jnd is around 0.5% for complex tones containing harmonics below the 7th, it rises above 2.5% when the lowest harmonic in the signal is the 13th or higher. Furthermore, Houtsma and Smurzynski demonstrated that the two levels of performance differ with regard to their sensitivity to the phase relations of the harmonics. When the harmonics are added in Schröder phase, one obtains a stimulus with the same spectral information, but with a temporally less differentiated structure. In this condition, the jnd remains the same for stimuli with lower harmonics, but it increases to 4%–5% when only higher harmonics are included. These results were

backed by Shackleton and Carlyon (1994) and by Carlyon and Shackleton (1994). They defined the resolvability of a harmonic complex in terms of the number of harmonics interacting within the same auditory filter as given by its 10-dB-down bandwidth, and they found that pitch matching for bandpass filtered harmonic complexes is not sensitive to phase relations of the harmonics when fewer than two harmonics interact (these are then said to be resolvable), while being sensitive to changes in phase relations when the number of interacting harmonics exceeds 3.25 (i.e., they are not resolvable). This would support the idea of two different pitch mechanisms working on resolvable and unresolvable harmonics, respectively. The latter mechanism would have to draw upon temporal regularities in the signal, while the former could also exploit spectral cues.

Psychophysical research on the “temporal” mechanism for the pitch of unresolved harmonics has mainly focused on the proof of its existence and on its dependence on the phase relations of the harmonics. Modeling of the temporal mechanism generally assumes some kind of autocorrelation algorithm to be at work. Kaernbach and Demany (1998) have demonstrated that first- and second-order temporal regularities are of quite different perceptual importance. This cannot easily be explained with any mechanism akin to an autocorrelation. However, due to the high cut frequency applied in their study (6 kHz), this conclusion may be limited to high-frequency regions. Furthermore, this high cut frequency entailed a degradation of the stimuli that let Kaernbach and Demany speak of “rattle” pitch rather than “musical” pitch. This constitutes an important drawback of their study as musicality is often considered a primary feature of pitch. It was the purpose of the present study to replicate and extend their

^{a)} Author to whom correspondence should be addressed. Electronic mail: christian@kaernbach.de

results at a lower cut frequency, and to test these sequences for their musicality.

Investigations into the temporal aspect of pitch perception have made use of a range of different stimuli. Apart from sinusoid complexes (e.g., Houtsma and Smurzynski, 1990), many experiments in psychoacoustics (e.g., Burns and Viemeister, 1981; Cariani and Delgutte, 1996a, b) and in physiology (see, e.g., Langner *et al.*, 1997) have applied sinusoidally amplitude-modulated (SAM) signals. Yost *et al.* (1998) investigated the temporal mechanisms involved in the analysis of iterated rippled noise (IRN, see also Yost, 1996). Kaernbach and Demany (1998) conducted their experiments with sequences of high-pass filtered clicks that were mixed with low-pass filtered noise. Periodic sequences of filtered clicks are equivalent to harmonic complexes added in cosine phase. This yields a temporally well-defined stimulus with a high peak factor and a low duty cycle. While the envelope of a SAM signal is higher than half of the maximum during 50% of a cycle, the temporal precision of a click train is only limited by the bandpass filtering done by the cochlea. Depending on the fundamental frequency of the train, one can obtain stimuli where, after cochlear filtering, the signal is higher than half of the maximum in less than 10% of a cycle. Furthermore, the interpretation of this stimulus as a series of filtered clicks admits aperiodic extensions of the paradigm where the stimulus conveys certain types of temporal regularities and excludes others. The experiments in the present article make use of these outlined advantages in different ways.

The first experiment reported in the present article is a variation of experiment III by Houtsma and Smurzynski (1990), using sequences of high-pass filtered clicks presented together with low-pass filtered noise. The jnd's reported by Houtsma and Smurzynski for complex tones with unresolvable harmonics and by Hoekstra (1979) for pulse trains are much larger than those they found for resolvable stimuli. Both studies used stimuli which cover only a small part of the cochlea. High-pass filtered clicks can offer more temporal information, i.e., on a larger part of the cochlea. Furthermore, there is no need to embed the sequences in pink noise which will disturb the signal portion of the stimulus. It should be sufficient to present low-pass filtered noise masking the region of possible distortion products. The aim of our experiment was to determine whether this optimized stimulus would have a lower jnd when containing only unresolvable harmonics, while still exhibiting the two-plateau jnd function. We would then be able to identify stimuli containing a maximum of temporal information but no resolvable harmonics, and use such stimuli to test further qualities of the elicited temporal pitch.

The next two experiments tested the musicality of these optimized temporal click sequences. Kaernbach and Demany reported informally that their stimuli did not convey musical information. However, this was not tested systematically. One reason for the lack of musical quality might be that the cut frequency applied by Kaernbach and Demany (6 kHz) was much higher than necessary to exclude resolvable information. On the basis of the data found in the first experiment, we test the ability to discriminate and classify musical inter-

vals constructed with periodic click sequences which are optimized with regard to their temporal properties.

Finally, a fourth experiment devised aperiodic click sequences that were similar to those applied by Kaernbach and Demany (1998). While their results could be argued to have validity only for frequency regions above 6 kHz, and that different mechanisms might be at work for frequencies which admit phase locking (i.e., below 4 kHz; see Rose *et al.*, 1967), the present study employs a cut frequency of 2 kHz that is well below this border. Once the applicability of their findings for these lower frequency regions was asserted, further conditions of the experiment using monaural and binaurally distributed irregular click sequences were devised to learn about the influences of such variations on the pitch percept.

II. EXPERIMENT 1: TWO LEVELS OF PITCH ANALYSIS

The jnd levels for temporal pitch analysis of 2% found by Hoekstra (1979) and of 2.5% reported by Houtsma and Smurzynski (1990) show that the temporal mechanism for unresolved harmonics is much less accurate than the mechanism for resolved harmonics. It should be noted that this performance was found for the isolated temporal mechanism, whereas the better performance for lower harmonics could be due to the cooperation of spectral and temporal mechanisms. It would nevertheless be highly interesting to find temporally defined stimuli that admit smaller jnd's.

In order to employ a temporally very accurate signal, periodic sequences of high-pass filtered clicks were devised which had characteristics similar to the sine-tone stimuli employed by Houtsma and Smurzynski or to the filtered click sequences used by Hoekstra. In contrast to those, however, they were not limited to a restricted frequency region either by a band-pass filter (Hoekstra) or by taking only a limited number of harmonics (Houtsma and Smurzynski) but contained all temporal information available up to the Nyquist frequency. It was anticipated that this would improve temporal pitch analysis. Furthermore, a different masker was used for the power spectrum below the signal. While Houtsma and Smurzynski had employed pink noise which has a certain intensity at the spectral location of the relevant signal components, in the present study white noise was used which was low-pass filtered such that its spectrum did not interfere with the signal, whereby a better signal-to-noise-ratio was obtained.

A. Stimuli and procedure

Periodic click sequences like those used in this experiment have a fundamental frequency equal to the number of clicks per second, and they contain all harmonics of the fundamental frequency with equal amplitude. In order to vary the lowest harmonic of the sequence, a high-pass filter was applied. The filter transition followed a logistic function containing several harmonics to avoid enhanced discriminability due to the appearance or disappearance of single harmonics at the cut frequency. The width was set to $\frac{1}{3}$ octave centered on the filter frequency, within which the amplitude increased from 10% to 90%. In order to ensure that the auditory power

spectrum would not contain resolved lower-order components arising from cochlear nonlinearity (Plomp, 1976), the signal was mixed with white noise. The noise was low-pass filtered at the same filter frequency as the sequence, using the inverted filter function such that the entire signal had a flat spectrum.

In the experiments, stimuli with fundamental frequencies of 100, 150, and 250 Hz were used. For each of these, the jnd was measured in 11 conditions, i.e., at eight different filter frequencies for filtered sequences mixed with noise as described above, at two filter frequencies for sequences without added noise, and for a pure tone of the fundamental frequency. The filter frequencies in the noise conditions were 500 Hz and 1, 1.5, 2, 2.5, 3, 4.5, and 6 kHz. In the two no-noise conditions, the sequences were filtered at 1 and 4.5 kHz, yielding signals with harmonics above the 10th and 45th, respectively, for a base frequency of 100 Hz, above the 6th and 30th for 150 Hz, and above the 4th and the 18th, for the base frequency of 250 Hz.

The experimental paradigm was an unforced weighted up-down adaptive procedure as described by Kaernbach (2001). On each trial, the subject was presented with a pair of stimuli A, B, which were presented twice sequentially in random order, i.e., either ABAB or BABA. The repetition aimed at eliminating the influence of previous trials. The task of the subject was to indicate via keyboard whether the last stimulus of the trial was the higher or the lower one, or whether he/she was not sure. Subjects were allowed to have a trial be repeated to them once. The subject had infinite time to answer and was given visual feedback on the correctness of his/her response, whereupon the next trial began.

Of the stimuli presented in a trial, one was constructed at a frequency f randomized with equal probability within $\pm 5\%$ of the base frequency f_0 , while the other had a frequency of $(1+p)f$, where p was varied adaptively. For convenience, a logarithmic scale was used for the adaptive procedure. A given difference level of d dB on this scale translates into p according to $p = 10^{d/10}$. The initial difference level d_0 for the runs was set to $d_0 = -12$ dB (approx. 6.3% of the base frequency). The initial step size was $\Delta d = 2$ dB. On a correct answer, the difference level was decreased by one step. If the answer was “unsure,” the difference level was increased by one step. After a wrong answer, it was increased by three steps. This leads to the point of 75% correct answers (Kaernbach, 2001). The maximum difference level admitted was -6 dB. The step size was halved after the third and after the fifth reversals. A run continued until the 16th reversal was reached, and the 75% threshold was then determined as the mean of all differences after the fourth reversal.

The experiment was run in a sound-proof both. The stimuli were digitally generated at 44.1 kHz and presented via electrostatic headphones at 60 dB SPL. The stimuli were presented diotically, with a length of 700 ms and a gap of 240 ms between each two. Where it was added, the noise started 240 ms before the first sequence and ended 240 ms after the fourth, including a linear on- and offset ramp of 150 ms. The pure tone stimuli had a ramp of 50 ms.

The experiment was divided into blocks comprising 11 runs, one run of each condition. The conditions were pre-

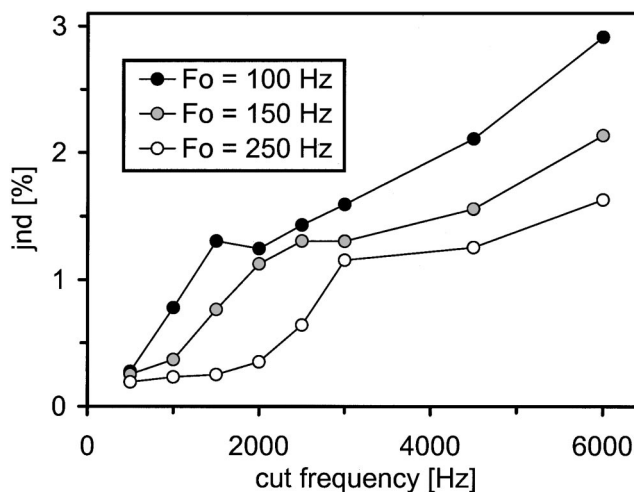


FIG. 1. Average jnd for the pitch of periodic click sequences in percent of the base frequency as a function of the filter frequency for three different fundamental frequencies. The two-plateau type of the jnd functions is clearly visible for all three functions, yet the exact position of the transition depends on the fundamental frequency.

sented alternately ordered per block, either in the order “mask,” “no mask,” “sine,” or vice versa. The order of the two “no mask” conditions as well as of the eight “mask” conditions was randomized. Ten subjects between 19 and 26 years of age took part, all of whom were students either majoring in music or having at least amateur musical experience of some years standing. For all but one, the jnd was measured six times in each condition; for one subject, the experiment was run three times. One subject was the second author; the other subjects were paid for their participation.

B. Results and discussion

Figure 1 shows the resulting mean jnd of the noise conditions in percent of the base frequency as a function of the filter frequency. Notably, the curves of the three base frequencies start at a similar jnd of below 0.3%, namely 0.19% for 250 Hz, 0.25% for 200 Hz, and 0.27% for 100 Hz, with standard deviations of means between 0.02 and 0.03. The three functions then climb with different slopes and level off at a similar jnd of just around 1.25%. The 100-Hz function does so at the filter frequency of 1.5 kHz with a jnd of 1.3%, the 150-Hz function at a filter frequency of 2 kHz with a jnd of 1.1%, and the 250-Hz curve levels off at a filter frequency of 3 kHz with a jnd of 1.15%. Similarly, the standard deviations of means rise to values around 0.1 to 0.2, with the exception of 0.3 for 100 Hz at a cut frequency of 6 kHz. When plotting the jnd as a function of the lowest harmonic rather than the absolute filter frequency, the jnd rises concurrently for all three base frequencies around the lowest harmonic of 10 (see Fig. 2). This verifies the dependence of the transition on the lowest harmonic contained in the signal.

The results for stimuli with resolvable harmonics are compatible with those reported by Houtsman and Smurzynski, who had found a jnd of 0.4 Hz for the base frequency of 200 Hz, which equals 0.2%. Also, their jnd function sloped between the lowest harmonic of 7 up to 13. A major difference between the results is the level at which the curves

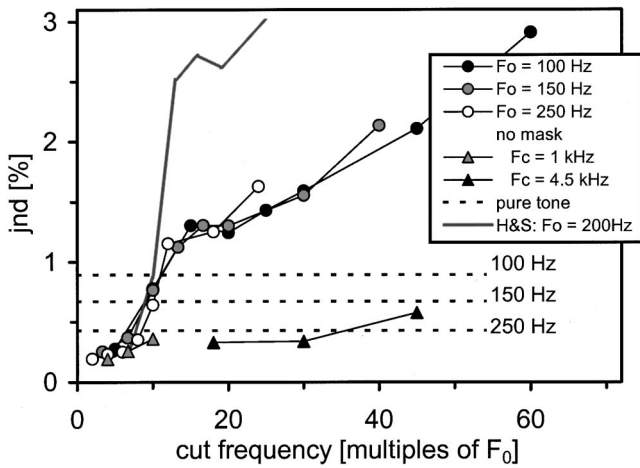


FIG. 2. Average jnd for the pitch of periodic click sequences in percent of the base frequency as a function of the lowest harmonic of the stimulus. As a function of harmonic number, the three jnd functions coincide. For comparison: jnd for complex tones of 11 harmonics of 200 Hz in sine phase as reported by Houtsma and Smurzynski (“H&S”). Also shown are the average jnd for single pure tones at the base frequencies (“pure”; from top to bottom: 100, 150, and 250 Hz) and the jnd for regular click sequences without masking noise (“no mask”); the three left-hand entries arise from the cut frequency of 1 kHz, the right-hand entries from 4.5 kHz.

saturate, as can be seen in Fig. 2. The jnd for the click sequences is considerably lower, which can be read as a direct indicator for the high temporal quality of the signal.

Figure 2 also depicts the jnd obtained for the “no mask” conditions and for the sine tones. Without masking noise, the jnd did not exceed 0.5% (with standard deviations of means ranging from 0.02 to 0.04), and there is no indication of a significant drop in performance between the two filter frequencies tested other than can be attributed to the decrease in signal. This emphasizes the importance of masking the lower part of the spectrum and reinforces the notion that the non-linear ascent of the curves in the noise condition is not attributable to the degradation of the signal, but to a qualitative change in processing. Without masking, the spectrally resolvable distortion products prevent this transition from occurring.

The jnd for the pure tone conditions are almost uniformly higher than those for the click sequences without noise, and they are also higher than the jnd for the lowest filter frequencies in the noise condition. The lowest pure tone jnd measured was 0.43% for 250 Hz, the others lay at 0.67% for 150 Hz and 0.89% for 100 Hz (with a standard deviation of means of 0.04, 0.09, and 0.07, respectively). This result comes as no surprise, since pure tones do not contain as much information as complex tones and, consequently, as click sequences conveying information within the dominant region of pitch perception (Ritsma, 1967).

An important detail to remark about the plateau in Fig. 2 is that it does not have a zero slope, but does slowly increase in a manner not unlike an exponential curve. Cullen and Long (1986) had similarly found a rate jnd increment from 4.5% up to 15% for a lowest harmonic of $N=13$ to $N=100$ (equivalent to a cutoff frequency of 10 kHz for a base frequency of 100 Hz). Since in experiment 1 the click sequences were not bandpass filtered, all harmonics up to the

Nyquist frequency were contained in the signal. Consequently, the number of harmonics gradually decreased as the filter frequency went up. However, the jnd cannot merely be a function of the number of harmonics, as Fig. 1 shows the opposite way, with 100 Hz showing the highest jnd, even though the stimulus contains the most harmonics when compared to the other two at a fixed filter frequency. Another hypothesis is that it could depend on the size of the cochlear region that is stimulated by the signal. In that case, all three curves should feature the same jnd for a fixed filter frequency, independent of the fundamental frequency. This does seem to be the case for 100 and 150 Hz at filter frequencies from 2 up to 3 kHz, but a separate explanation would have to be provided for the divergent progression of the jnds above 3 kHz.

Evidence that the two levels of performance do arise from two different mechanisms is revealed by examining the correlations of individual performances within and across the levels of the noise conditions. When correlating individual performances for any two noise conditions which both include harmonics below the 10th, we find an average value of 0.8, with a standard deviation of 0.12. Likewise, we find an average correlation of 0.69, with a standard deviation of 0.18, when pairwise correlating individual performances for conditions which include harmonics only above the 10th. This means that the “rankings” of the subjects are quite consistent within each of the two levels. In contrast, correlating performances across the transition yields an average correlation of 0.35, with a standard deviation of 0.2, i.e., the rankings are not stable between the two levels. This is a strong indication that the mechanisms involved in pitch analysis of stimuli with and of those without spectrally resolvable components differ.

III. EXPERIMENTS 2a AND b: THE MUSICALITY OF TEMPORAL PITCH

Musicality is an important feature of pitch perception. Several auditory stimuli convey a perception of “high” and “low” which does, however, not compare to pitch. It is often suggested that the identification of musical intervals may serve as a criterion for pitch perception. We wanted to test how far the stimuli used in the first experiment can convey musical information. It is obvious that by eliminating the resolvable harmonic information, one reduces the quality of pitch of the stimuli which are applied. This does not degrade the importance of temporal cues, as with natural stimuli in general both spectral and temporal analysis are possible and will probably both contribute to the perception of a high-quality and musical pitch. It would nevertheless be interesting to see whether to some degree musical quality remains within a purely temporal signal, or whether the resulting stimuli convey but a “rattle pitch” (Plomp, 1976, Chap. 7). Perhaps musicality does not constitute an all-or-none phenomenon but a quality with all kinds of gradations that depend on the amount of information left in the stimulus. A critical test for this view would be that those subjects that

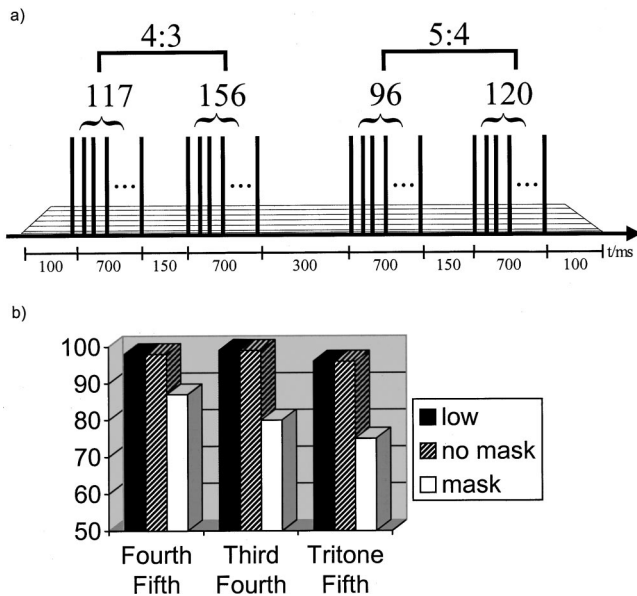


FIG. 3. (a) Schematic example of an interval discrimination trial. It consists of four click sequences embedded in noise. The first two sequences have a frequency ratio of 156 clicks to 117 clicks, equaling $\frac{4}{3}$ or a fourth, and the following two form a major third. (b) Average performance on musical interval discrimination for complexes comprising low harmonics (“low”), high-pass filtered click sequences without noise masker (“no mask”) and with noise masker (“mask”). For each stimulus, discrimination was tested for a major third against a fourth, a fourth against a fifth, and a fourth against a tritone. Performance is shown in percent of total trials.

show superior pitch processing for unresolvable harmonics in terms of their jnd will also have superior musical interval recognition.

A. Experiment 2a: Stimuli and procedure

In this experiment, discrimination tasks were run for three pairs of intervals: For a major third (frequency ratio of $\frac{5}{4}$) versus a fourth (frequency ratio of $\frac{4}{3}$), fourth against fifth (frequency ratio of $\frac{3}{2}$), and a tritone (frequency ratio of $\frac{45}{32}$) versus fifth. The target interval was the smaller one of both in all three cases. The discrimination of a fourth against a fifth was assumed to be the easiest. The ratio of the two frequency ratios is $\frac{8}{9}$, that is, the difference between the two intervals is one full tone. The other two discrimination tasks are nominally of the same difficulty, since in both cases the ratio of the frequency ratios is $\frac{15}{16}$, i.e., the difference between the intervals is a semitone in both cases. Yet, the last comparison (tritone versus fifth) should be more difficult, for two reasons. First, following Weber’s law, it should be easier to judge a certain interval difference for smaller intervals than for larger ones. Second, while the comparison of a major third to a fourth comprises intervals that are quite familiar in normal musical contexts, the tritone represents an interval that will be not equally familiar to the subjects.

Each pair of intervals was tested for in three stimulus conditions: first, for periodic click sequences with low-pass filtered noise, second, for click sequences without a noise masker, and finally, for complex tones comprising the fundamental frequency itself as well as the next five harmonics (i.e., 1f,...,6f) added in sine phase. Figure 3(b) shows an example trial for click sequences with noise.

On each trial, two intervals were sequentially presented in random order, the lower tone preceding the higher one in each interval. Each tone was 700 ms long (the complex tones including a linear on-and offset ramp of 50 ms), with gaps of 150 ms between each two tones belonging to the same interval and a gap of 300 ms between the two intervals. In the noise condition, the noise started 100 ms before the first tone and ended 100 ms after the last, including a linear on- and offset ramp of 50 ms. Trials were grouped into blocks of 50 which were presented in random order in sets containing one block of each condition.

The frequency of the higher tone of each interval was randomized with equal probability within the range from 150 to 200 Hz; the lower tones were then computed accordingly. In order to guarantee that the signal was only temporally resolvable, the lowest harmonic was set to 15 in accordance with the results described in Sec. II. Given the maximum of 200 Hz for the base frequency, this condition equals a filter frequency of 3 kHz. Condition was made that the higher and the lower tones of two intervals tested for in one trial had to differ by 5% to ensure that subjects could not deduce the correct interval by comparison of either the higher or lower notes of the intervals.

Six subjects participated in the experiment, all of whom had also taken part in the experiment described in Sec. II. One subject accomplished five sets of blocks, all other subjects accomplished six sets. For each subject, the first set was discarded.

B. Experiment 2a: Results and discussion

Figure 3(b) shows the average discrimination rates for the three conditions and the three discrimination tasks. As predicted, performance is best for the fourth/fifth discrimination, and is lowest for the tritone/fifth discrimination. This holds for all three conditions. The two conditions without a noise masker (low partials versus high partials) practically yield the same results, i.e., 99% for the fourth/fifth discrimination (with a standard deviation of means of 0.3), 98% for discrimination of a third versus a fourth (with a standard deviation of means of 0.7 for low partials and 0.4 for high partials), and 96% for the tritone/fifth discrimination (with a standard deviation of 1.0). Although performance degrades for the noise conditions, performance clearly remains above chance level. Discrimination is 80% for fourth/fifth (with a standard deviation of means of 2.3), 87% for third/fourth (standard deviation of means: 1.8), and 75% for tritone/fifth (standard deviation of means: 1.7). Introspection indicates actual musical interval identification: Some subjects report singing the intervals in order to discriminate between them.

As subjects were matched, it was possible to correlate levels of performance between the jnd of temporal pitch and the interval discrimination. Figure 4 shows the interval discrimination performance for the tritone/fifth discrimination as a function of the corresponding jnd from experiment 1. There is a clear correlation of -0.75 to the effect that interval discrimination is worse the higher the jnd. This reinforces the view that for our stimuli the musicality of pitch is a gradual function of the reliability of recognition.

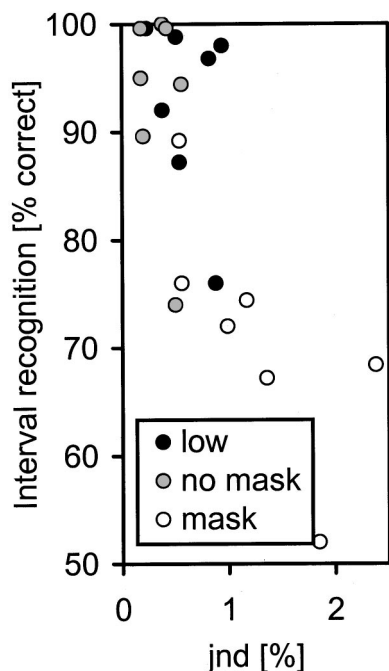


FIG. 4. Performance for musical interval discrimination of individual subjects (experiment 2) plotted against the jnd of pitch analysis (experiment 1). The labels “low,” “no mask,” and “mask” refer to the same conditions as in Fig. 3. The individual performance on interval discrimination shows to be correlated with the jnd, especially in the “mask” condition.

C. Experiment 2b: Stimuli and procedure

Even though introspection indicates that in experiment 2a judgments were based on the musical identification of the intervals, the decisions could theoretically have merely been based on a comparison of the interval sizes. To eliminate this possibility, in the second musicality experiment, subjects had to classify intervals directly. On each trial, they were presented with two stimuli, which had an evenly randomized frequency ratio between 1.0 and 2.0. They had to classify the interval realized by the two stimuli by choosing the most adequate musical interval on a labeled computer keyboard. The task involved having to classify ambiguous intervals, as the actual frequency ratio of the stimuli could lie anywhere between two consecutive musical intervals. Consequently, the subjects received no feedback on the accuracy of their classification. The stimuli could be replayed at leisure via two keys on a computer keyboard, and subjects were encouraged to take as much time as they thought necessary to arrive at the best possible classification.

Three conditions were devised: In a “musical” control condition, the stimuli were unfiltered periodic click sequences in the octave from 80 to 160 Hz. These sequences include all harmonics of the base frequency, resolvable as well as unresolvable. The “temporal” condition used sequences constructed within the same octave, but containing only unresolvable harmonics. As in the previous experiment, this meant eliminating all spectral components below the 15th harmonic. Both stimuli of a trial were high-pass filtered at the adequate filter frequency for the higher of the two base frequencies, i.e., between 1.2 and 2.4 kHz, and had low-pass filtered white noise added. The “infra-pitch” condition was explicitly designed not to be musical. It used infra-pitch click

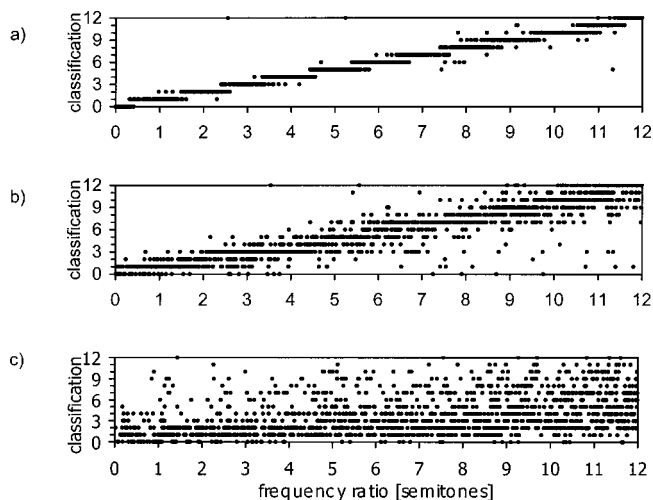


FIG. 5. Classifications as a function of the actual stimuli frequency ratio. (a) Musical condition. The classifications cluster tightly around the main diagonal. (b) Temporal condition. There is some scatter, but the clustering is still recognizable. (c) Infra-pitch condition. The classifications retain little structure.

sequences in the range from 8 to 16 Hz (Pressnitzer *et al.*, 1999), i.e., containing between eight and 16 evenly spaced clicks per second.

Due to the continuous range of the stimulus frequency ratios, a “perfect” classification was not possible. The best classification would map onto each musical interval the range of frequency ratios closest to this respective interval. It was anticipated that, for the first two conditions, subjects would get close to this level of performance, while basically having to guess in the third condition.

For each trial, a frequency ratio between 1.0 and 2.0 was randomly chosen. The lower frequency was then randomly set within the octave of possible frequencies (80 to 160 Hz, or 8 to 16 Hz) with the restriction that the higher frequency would also lie within the octave. This entailed that a trial with a frequency ratio of 2.0 would always be made up of the lower and the upper boundary of the respective octave.

For the first two conditions, each stimulus had a length of 1 s. In the filtered condition, the actual sequence was just 900 m long with a noise on- and offset of 50 ms each. In third condition, the length of the stimuli was randomized to have a length between 1 and 1.4 s in order to make sure that the classification would not rely on counting the number of clicks in the stimuli. All stimuli had linear on- and offset amplitude ramps of 150 m.

Three subjects with professional musical background took part in the experiment. Two of them completed 15 blocks in each condition, one completed 5 blocks. Each block comprised 50 trials. The first block of each condition and subject was discarded.

D. Experiment 2b: Results

As expected, classification was reported to be easy for the musical condition, more difficult, yet readily performable, for the temporal condition, and impossible for the infra-pitch condition. This qualitative impression was matched by the data. Figures 5(a)–(c) plots the classification as a func-

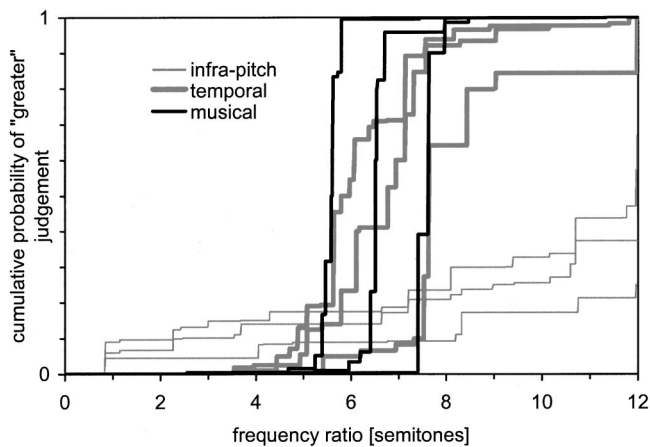


FIG. 6. The cumulative psychometric function for the judgment that the frequency ratio is higher than six, seven, and eight semitones, respectively, plotted for all three conditions. While there is a visible transition for the control condition as well as the temporal condition, the curve fails to reach 50% in the infra-pitch condition.

tion of the actual frequency ratio. Classification for the musical condition is nearly ideal, the correlation between actual frequency ratio and answer being 0.99. The classifications for the temporal condition group around the main diagonal in a similar manner. Even though there is visibly more scatter, the correlation is still 0.89. This is obviously in contrast to the result of the infra-pitch condition, in which classifications are widely spread. Here, the correlation is as low as 0.4.

A more accurate picture can be drawn by looking at the transitions of judgment between consecutive semitones. This is shown in Fig. 6, which plots the cumulative judgment transitions from 5th to 6th, from 6th to 7th, and from 7th to 8th, for all three conditions, as obtained using a pool-adjacent-violators algorithm. While the musical condition as well as the temporal condition both show a clear shift in judgment, the infra-pitch condition remains below 50% until the end of the abscissa. In the musical condition, the transitions were found to lie almost perfectly in the middle between semitones. For the temporal condition, there was an overall bias of the mean towards the higher semitone, i.e., the subject would tend to prefer the lower one. For the infra-pitch condition, transitions as such cannot be observed.

In conclusion, the three conditions reflect a gradual degradation of the ability of the subjects to classify the intervals in the order musical condition, temporal condition, and infra-pitch condition. However, the performance for the temporal condition is noticeably closer to that for the musical condition than to the performance for the infra-pitch condition. Consequently, if musicality were rigidly to be thought of as a quality either present or not in a pitch percept, instead of as a quality which can deteriorate gradually, the boundary would have to be set between the musical and the temporal conditions on the one side, and the infra-pitch condition on the other.

IV. EXPERIMENT 3: FINDING TEMPORAL REGULARITIES IN APERIODIC CLICK SEQUENCES

One of the major advantages of using sequences of filtered clicks instead of complexes of harmonically related

components is the possibility to construct sequences that are not periodic but show some more complex kind of temporal regularity. Using such aperiodic sequences, Kaernbach and Demany (1998) demonstrated that the temporal mechanism involved in the pitch of unresolved harmonics can deal significantly better with first-order regularities (i.e., relating to the distances between a click and its direct successor) than with second-order regularities (relating to the distance between nonsuccessive clicks). A problem of their study was the rather high cut frequency of 6 kHz. Apart from yielding conceptually poor stimuli which reportedly contained but a “rattle pitch,” this frequency is beyond what generally is believed to be the limit to phase locking in auditory nerves (e.g., see Rose *et al.*, 1967). It is not certain how far this should be an important limit to a temporal mechanism that is supposed to operate on envelopes rather than on the fine structure of the signal. However, in order to obtain results that can be considered to be relevant to all kinds of temporal processing, it would be favorable to choose the filter frequency below 4 kHz.

Therefore, it was the aim of this experiment to replicate the results of Kaernbach and Demany at a cut frequency of 2 kHz. Furthermore, the approach of Kaernbach and Demany was taken a step further by splitting click sequences onto both ears, whereby a dichotic stimulus was obtained which had binaurally integrated interclick interval (ICI) statistics different from the ICI statistics for its two single, monaural parts. It was then possible to examine whether subjects would deal with these stimuli mainly according to the monaural characteristics or whether the statistics of the binaurally integrated stream would prevail.

A. Stimuli and procedure

In the first two conditions [“complete” sequences, see Fig. 7(a)], which mainly aimed at verifying the results obtained by Kaernbach and Demany, subjects had to discriminate diotic *kxx*- and *abx*-sequences from random (*x*-) sequences. Since it was also the purpose of these control conditions to see whether the results would be confirmed for a filter frequency considerably below 4 kHz, the spacing of the sequences had to be chosen differently: *k* was set to 7.5 ms, and *a + b* was set to 15 ms. Accordingly, *x* was randomized within [0, 15] ms. This equals an average of approximately 133 clicks per second, or a fundamental frequency of 133 Hz for the sequence, thereby admitting a filter frequency of 2 kHz. Filtering and masking with filtered white noise were done in the same manner as has been described in Sec. II A. The same noise was used in all three conditions. For the random sequences, a restriction devised by Kaernbach and Demany was adopted which ensured that the maximum of consecutive intervals either all within the higher or all within the lower half of the interval distribution would be comparable to the target sequence. For a *kxx*-sequence this number cannot exceed 2, and for *abx* the limit is 3 (see Kaernbach and Demany, 1998).

In the following two conditions [“semi” sequences, see Fig. 7(b)] the stimuli were generated according to the same rules as before, with the difference that every second click was left out. The semi-sequences show complementary char-

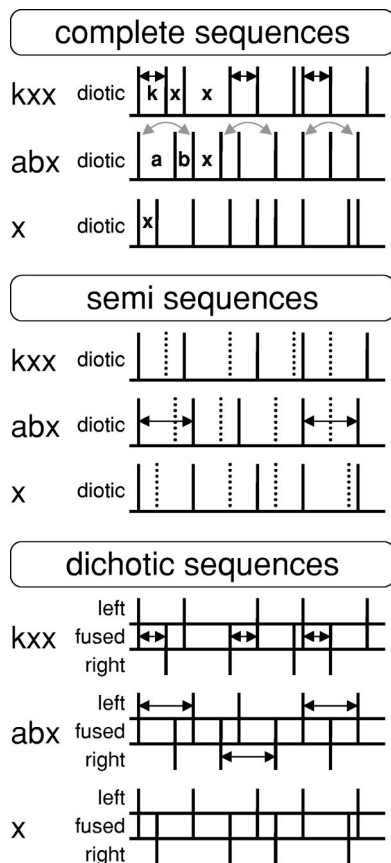


FIG. 7. (a) Schematic representation of kxx , abx , and random (x) click sequences as devised by Kaernbach and Demany (1998). The abscissa denotes course of time, vertical lines represent clicks. In a kxx sequence, the first (“ k ”) of each three intervals between two clicks is set to a constant time $k = 7.5$ ms, while the following two (“ x ”) are randomized within $[0, 15]$ ms. An abx sequence is constructed such that the first (“ a ”) of each three intervals is taken randomly from a uniform distribution $[0, 15]$ ms, and the second interval (“ b ”) such that $a + b$ add up to a second-order interval of constant time of 15 ms with a click in between inserted at a random position. The third interval (“ x ”) is randomly chosen from $[0, 15]$ ms. A random sequence contains intervals that are randomized within $[0, 15]$ ms. (b) Diagram of the modified “semi”-click sequences. In these sequences, every second click is left out. These are represented by the dotted lines. As can be seen, it is now with abx sequences that first-order intervals occur. If perception is based on first-order intervals, abx should now be easier to detect than kxx . (c) Dichotic kxx , abx , and random click sequences. The clicks are presented alternately to the left and to the right ear. Listening to one of the monaural streams would admit the detection of first-order intervals in abx , while listening to the binaurally fused stream would help to detect the first-order intervals in kxx .

acteristics to the complete sequences: While the abx semi sequences have a distinct first-order ICI peak at $a + b$ ms, the first-order ICI statistics of the kxx semi sequences is comparable to that of the random sequences. As a consequence, exactly the opposite discrimination performance was expected in these conditions.

Finally, in a fifth and sixth condition [“dichotic” sequences, see Fig. 7(c)], the stimuli were each split into two complementary semi sequences, which were presented dichotically, one to the left ear and one to the right. In other words, the clicks are presented to the ears alternately. Looking at each monaural stream by itself, the stimuli cannot be distinguished from the semi-sequences [Fig. 7(b)], whereas the integrated stream of both ears reveals just an original,

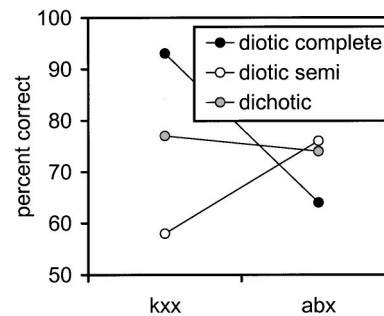


FIG. 8. Average performance for discrimination of kxx and abx against random sequences. The sequences were either presented diotically (“complete”), diotically with every second click left out (“semi”), or alternated between both ears (“dichotic”). Results are plotted as percentage correct of total of trials. While kxx can be discriminated perfectly in the complete condition and abx cannot, an inverse performance pattern at a somewhat lower overall performance level is to be observed for the semi condition. In the dichotic condition, abx can be discriminated as in the semi condition, while kxx can be discriminated at a level somewhat below the complete condition, but significantly above the semi condition.

“complete” sequence [Fig. 7(a)]. So if the discrimination was accomplished by means of the binaurally integrated stream, the outcome should resemble the results from the “complete” conditions. If, on the other hand, the separate monaural streams were the basis for the discrimination, the results should be similar to those of the “semi” conditions.

Each trial consisted of a target sequence and an adequate (i.e., complete, semi or dichotic) random sequence, which were presented sequentially in random order for a duration of 750 ms each, with a gap of 250 ms in between. The low-pass filtered noise began 250 ms before the first sequence and ended 250 ms after the second, including a linear on- and offset of 100 ms. The subject was allowed to have the trial repeated once. He/she was to indicate via keyboard whether the target sequence had been the first or the second sequence. There was no time limit imposed. Upon answer, visual feedback was given. If the answer was incorrect, the subject was allowed to have the trial played once more. Then the next trial began.

Trials were grouped into blocks of 80, of which the first 40 were declared to be test trials and were excluded from evaluation. The blocks for the different conditions were presented in random order. A set of six blocks, one block for each condition, had to be completed before a new set began. The subjects were given information about which condition the current block belonged to, the conditions being neutrally named “A,” “B” to “F” condition. Five subjects took part in the experiment, all being students between 20 to 24 years of age. None had previous experience with psychoacoustic tasks. All subjects accomplished 18 sets of blocks of which the first 3 were excluded from evaluation.

B. Results and discussion

The results for the discrimination in the six different conditions are shown in Fig. 8. Performance for the complete conditions is comparable to what Kaernbach and Demany had found: The kxx sequences were discriminated nearly perfectly, while the abx sequences were discriminated with an average chance of 64%.

For the semi sequences, an inverted performance pattern is observed, as had been anticipated. The overall lower performance as compared to the complete condition is most likely due to the lower number of ICIs in these stimuli, i.e., it merely reflects a quantitative difference of the temporal information contained in the two types of sequences, not a qualitative difference.

Finally, for the dichotic sequences, the result is twofold: On the one hand, discrimination of *abx* can be done with the same performance as for the *abx* semi sequences. This necessarily implies that the task can be solved by means of the monaurally separated streams. It appears that only one of these streams was evaluated at each trial, as the performance is very close to that of a single semi sequence. On the other hand, the performance for the dichotic *kxx* sequences remains significantly above that for the semi *kxx* sequences and below that for the complete *kxx* sequences. This entails that the subjects were able to draw upon cues from the binaural stream, but could not to fuse the streams completely and obtain the performance level of the first condition. It would seem that the mechanism underlying discrimination is not exclusively committed to either binaural or monaural processing. However, the degradation of performance for the dichotic *kxx* sequences as compared to the complete *kxx* sequences suggests that monaural processing is more robust than binaural processing. Assuming binaural processing to be more robust, one would have expected such a degradation to occur for the dichotic *abx* sequences. Similarly, a prevalence of monaural processing has recently been reported by Carlyon *et al.* (2001), who found no binaural processing in streams with monaural regularities. Presumably, binaural processing will only work on streams with very weak or no monaural cues, as in the present experiment with the dichotic *kxx* sequences. It remains to be seen in future studies how far this prevalence might be influenced by attentional mechanisms (e.g., by directing the subject's attention towards monaural or binaural listening).

To control for the possibility of physical interaural cross-talk being the cause of the high performance on dichotic *kxx* sequences, these sequences were recorded with an artificial head. Cross-talk was found to occur at 24 dB below presentation level. The monaural halves of the recorded dichotic *kxx* sequences could not be discriminated above chance from random sequences. If there had been additional information provided by cross-talk of clicks, this performance should have been comparable to that for dichotic *kxx* sequences.

V. CONCLUSIONS

High-pass filtered and low-pass masked click sequences are an excellent tool to investigate the temporal mechanisms involved in auditory pitch perception. Their high temporal definition produces behavioral performances that are superior to those achieved with other stimuli (see experiment 1). At the same time, the difference between the performance for low and high harmonics remains sufficiently distinct to tell clearly where the domain of purely temporal pitch processing begins. The latter should not be considered as a separate phenomenon but as a mechanism that contributes to pitch

perception over the entire region of the spectrum and can be sounded in isolation only for unresolved harmonics. It is obvious that the performance is lower for signals that sound this isolated temporal mechanism than for signals where it can be seconded by spectral cues. The classification of intervals constructed with unresolvable stimuli functions at a level comparable to that reached for fully musical stimuli (see experiment 2b). Subjects with a low temporal pitch jnd discriminate musical intervals constructed with spectrally unresolvable stimuli better than subjects with a higher temporal jnd (experiment 2a). This supports the notion of a unified pitch percept that draws from several mechanisms, one of them being the temporal mechanism under study.

A subject matter of interest is the nonzero slope of the right-hand part of the two-plateau jnd functions, i.e., the jnd for the unresolved harmonics (Fig. 2). It would be highly elucidating if one could find a variation of the experiment which would yield really flat plateau functions. Experiment 1 leaves open the question whether this could be achieved at least for a certain range of fundamental frequencies and of filter frequencies by covarying the upper limit of bandpass filtered clicks with the lower limit. It would be most interesting to find that with a constant ratio of upper and lower cut frequency (i.e., with the signal directed to a basilar membrane segment of constant length) one would find a constant jnd independent of the lower cut frequency, i.e., of the spectral region. This would support the idea that the temporal mechanism does operate on the envelope of the excitation of the basilar membrane for specific frequency ranges, as the envelope for unresolvable complex tones does not depend on the harmonic number of the carrier frequency.

Beside their high temporal definition, filtered click sequences tackle the temporal mechanism involved in pitch perception in a very direct and straightforward way, as they "speak the language of the auditory nerve." Therefore, experiments can be designed that test specific hypotheses about this mechanism very directly. Experiment 3 is an example of the flexibility that is inherent to this kind of stimulation. The results for the "complete" sequences replicate those of experiment 3 of Kaernbach and Demany with a lower cut frequency. There seems to be no essential difference of temporal processing of first- and second-order regularities for spectral regions above and below 4 kHz. This supports the idea of an envelope analysis process that does not depend on the phase locking to the fine structure of the carrier. This mechanism operates more easily on first-order than on second-order temporal regularities, a fact that should be considered seriously by theoretical modelers of this temporal mechanism. The binaural condition of experiment 3 reveals that both monaural and binaural processing of temporal pitch is possible. This parallels similar results from spectral pitch according to which it seems possible to integrate the spectra of both ears before applying a spectral pattern algorithm (Houtsma and Goldstein, 1972). This could be an indication for the existence of early interaction between temporal and spectral pitch mechanisms.

ACKNOWLEDGMENTS

The authors would like to thank Laurent Demany, Peter Cariani, and an anonymous reviewer for their helpful comments on earlier drafts of the manuscript. This research was supported by the DFG Grant No. KA 824/5-1.

- Burns, E. M., and Viemeister, N. F. (1981). "Played-again SAM: Further observations on the pitch of amplitude-modulated noise," *J. Acoust. Soc. Am.* **70**, 1655–1660.
- Cariani, P. A., and Delgutte, B. (1996a). "Neural correlates of the pitch of complex tones. I. Pitch and pitch salience," *J. Neurophysiol.* **76**, 1698–1716.
- Cariani, P. A., and Delgutte, B. (1996b). "Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase-invariance, pitch circularity, rate-pitch, and the dominance region of pitch," *J. Neurophysiol.* **76**, 1698–1716.
- Carlyon, R. P., and Shackleton, T. M. (1994). "Comparing the fundamental frequencies of resolved and unresolved harmonics: Evidence for two pitch mechanisms?" *J. Acoust. Soc. Am.* **95**, 3541–3554.
- Carlyon, R. P., Demany, L., and Deeks, J. (2001). "Temporal pitch perception and the binaural system," (to appear).
- Cullen, Jr., J. K., and Long, G. (1986). "Rate discrimination of high-pass-filtered pulse trains," *J. Acoust. Soc. Am.* **79**, 114–119.
- Hoekstra, A. (1979). "Frequency discrimination and frequency analysis in hearing," unpublished Ph.D. thesis, University of Groningen.
- Houtsma, A., and Goldstein, J. L. (1972). "The central origin of the pitch of complex tones: evidence from musical interval recognition," *J. Acoust. Soc. Am.* **51**, 520–529.
- Houtsma, A., and Smurzynski, J. (1990). "Pitch identification and discrimination for complex tones with many harmonics," *J. Acoust. Soc. Am.* **87**, 304–310.
- Kaernbach, C. (2001). "Adaptive threshold estimation with unforced-choice tasks," *Percept. Psychophys.* **63**(8) (in press).
- Kaernbach, C., and Demany, L. (1998). "Psychophysical evidence against the autocorrelation theory of auditory temporal processing," *J. Acoust. Soc. Am.* **104**, 2298–2306.
- Langner, G., Sams, M., Heil, P., and Schulze, H. (1997). Frequency and periodicity are represented in orthogonal maps in the human auditory cortex: evidence from magnetoencephalography," *J. Comp. Physiol., A* **181**, 665–676.
- Plomp, R. (1976). *Aspects of Tone Sensation* (Academic, London).
- Pressnitzer, D., Patterson, R. D., and Krumbholz, K. (1999). "The lower limit of melodic pitch for filtered harmonic complexes," *J. Acoust. Soc. Am.* **105**, 1152.
- Ritsma, R. J. (1967). "Frequencies dominant in the perception of the pitch of complex sounds," *J. Acoust. Soc. Am.* **42**, 191–198.
- Rose, J. E., Brugge, J. F., Anderson, D. J., and Hind, J. E. (1967). "Phase-locking responses to low-frequency tones in single auditory-nerve fibers of the squirrel monkey," *J. Neurophysiol.* **30**, 769–793.
- Schouten, J. F. (1940). "The residue and the mechanism of hearing," *Proc. K. Ned. Akad. Wet.* **43**, 991–999.
- Schouten, J. F. (1970). "The residue revisited," in *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G. F. Smoorenburg (Sijthoff, Leiden), pp. 41–54.
- Shackleton, T. M., and Carlyon, R. P. (1994). "The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination," *J. Acoust. Soc. Am.* **95**, 3529–3540.
- Yost, W. A. (1996). "Pitch of iterated rippled noise," *J. Acoust. Soc. Am.* **100**, 511–518.
- Yost, W. A., Patterson, R., and Sheft, S. (1998). "The role of the envelope in processing iterated rippled noise," *J. Acoust. Soc. Am.* **104**, 2349–2361.