Published in ACTA ACUSTICA UNITED WITH ACUSTICA, Vol. 94, Pgs. 594–603 (2008). [ISSN 1610-1928]
(The Journal of the European Acoustics Association (EAA). International Journal on Acoustics.)

Temporal resolution of hearing probed by bandwidth restriction

Milind N. Kunchur
Department of Physics and Astronomy
University of South Carolina, Columbia, SC 29208
Email: kunchur@sc.edu
(Dated: Received: 10 August 2007. Accepted: 21 May 2008)

This work investigates the temporal resolution of human hearing through its discrimination of the time constant of low-pass filtering applied to a periodic signal. While restricting the bandwidth affects both the amplitude spectrum and temporal definition of the signal, the direct amplitude changes in this experiment fall below their just noticeable differences. The discrimination therefore seems to be sensitive to phase in addition to spectral amplitude differences. An upperbound of $\tau \approx 5$ μs was obtained for the threshold time constant, showing that human temporal resolution extends down to time scales shorter than found in the past.

PACS numbers: 43.66.Mk, 43.66.Jh, 43.66.Ba.

I. INTRODUCTION

The temporal resolution and high-frequency audibility of human hearing are complex issues of both fundamental and practical significance. While the single-tone high-frequency threshold f_{max} for airborne stimuli is around 18 kHz in individuals with good hearing (Pumphrey, 1950; Hall, 2002), a much higher bandwidth and temporal acuity can play a role in the complete perception of the timbre of sound. Neural processing beyond the cochlea can permit extraction of temporal information at time scales τ shorter than the $1/2\pi f_{max}$=9 μs that would be nominally expected for a linear system. In binaural localization by interaural time difference, it is well known that differences in arrival times of order 10 μs are distinguishable (Henning, 1974; Nordmark, 1976). Monoaural experiments involving iterated ripple noise (IRN) and inter-pulse gaps have shown similar thresholds in temporal resolution (Krumbholz, 2003; Leshowitz, 1971). A similar sensitivity for temporal fine structure can be inferred from the discriminability of the virtual pitch of complex tones (Moore et al., 2006; Gockel et al., 2006). It also appears that the cochlea may sense ultrasonic stimulation if the latter manages to reach the cochlea in sufficient intensity, both when presented through the air (Henry and Fast, 1984; Ashihara *et al.,* 2006) but especially when presented through bone conduction (Corso, 1963; Deatherage *et al.,* 1954; Lenhardt *et al.,* 1991; Lenhardt, 1998). It has also been conjectured that such high level ultrasound may possibly change the perception of timbre when superimposed on audible harmonics (Oohashi *et al.,* 1991; Yoshikawa *et al.,* 1995). Additionally, restricting the bandwidth by low-pass filtering necessarily attenuates all frequencies to some extent, and hence spectral amplitude changes can never be avoided absolutely (even when $1/\tau \gg f_{max}$); how those amplitude changes affect timbre will depend on their magnitudes relative to the relevant just noticeable differences. For these reasons it can be expected that limiting the bandwidth of an audio signal by low-pass filtering may produce an audible change, even when the high-frequency cutoff (or equivalently $[2\pi\tau]^{-1}$) is well above f_{max}. The present work experimentally confirms this to be true, and at intensity levels and time constants much lower than suspected possible before.

Temporally disparate waveforms can be distinguished through two mechanisms. The first involves temporal discrimination through higher-level neural processing beyond the cochlea—such as through a possible mechanism that encodes and represents at higher levels (e.g., the inferior colliculus) the time intervals between peaks of phase-locked signals conveyed by the auditory nerve fibers (e.g., Meddis and Hewitt, 1991; Patterson *et al.,* 1995; Krumbholz *et al.,* 2003). The second mechanism for discriminating temporally disparate waveforms involves sensing differences in spectra from excitation patterns of the hair cells. To distinguish between the two mechanisms, the temporally disparate waveforms need to be isospectral in amplitude.

In experiments probing temporal resolution, a pair of stimuli are presented that differ in their temporal structure. As the temporal difference is progressively reduced, one finds the threshold for barely being able to discern a difference. In one experiment by Leshowitz (1971), listeners were presented with a single pulse or two narrower pulses (with the same total energy) separated by an interval Δt. The click and click-pair could be distinguished down to $\Delta t \approx 10$ μs. In this case, the two stimuli have differences in their amplitude spectra and their discernment was explained on this basis. Isospectral variants of this experiment were carried out by Ronken (1970) and later by Henning and Gaskell (1981) where one stimulus consisted of a short pulse followed by a taller one separated by an interval Δt. The second stimulus was a similar pair with the time order reversed and hence had the same amplitude spectrum. The shortest Δt for which these stimuli could be distinguished was about 200 μs. Another type of constant-amplitude-spectrum experiment involves the detection of gaps in noise (Plomp, 1964; Penner, 1977; Eddins *et al.,* 1992). In these the threshold for gap detection was of the order of 2 ms. The issue of determining temporal resolution while avoiding spectral cues was recently tackled (Yost *et al.,* 1996; Patterson and Datta, 1996; Krumbholz *et al.,* 2003) through the use of iterated rippled noise. The experiment of Krumbholz et al. (2003) showed that differences in delay between a masker and signal could be discerned down to 12.5 μs. In their work a masking paradigm was used to argue that spectral cues did not play a role in the discernment. Note that in all the previous cited experiments, the threshold Δt exceeded the nominal 9 μs.

The preceding summary of earlier work highlights the difficulty in unambiguously separating the effects of the time and frequency domains. The experiment presented in this paper provides a new window on this problem: Instead of invoking a masking paradigm, the issue of spectral cues is probed by comparing changes in sound pressure level and harmonic amplitudes to their just noticeable differences. Concerning the bandwidth standards for audio reproduction, the demonstrated audibility of a 5 μs time constant is significant regardless of which of the two domains is/are operative in the perception. The result shows that sampling rates in consumer digital audio are insufficient for complete fidelity.

II. METHODS

The experiment consists of presenting an approximately square-wave shaped complex tone, with a 7 kHz fundamental, through earphones with different degrees of low-pass filtering (i.e., with different time constants τ) and testing a listener's ability to distinguish this filtered tone from the unfiltered control tone (τ=0).

A. Apparatus

A significant potential bottleneck in a temporal-resolution experiment is the temporal-response speed of the equipment. Typically the apparatus consists of signal sources, a switching/gating method used to ramp the signals, an amplifier for driving the tranducer, and the transducer itself. In the present work, many different approaches were initally tried and abandoned, including using digital synthesis (with 24-bit/96-kHz sampling) for the production and ramping of signals. It was found that such a digital method had far too inadequate temporal definition for this purpose. So instead an analog signal generator (model 4001 manufactured by Global Specialties Instruments, Cheshire, Connecticutt) was used to produce a 7 kHz square waveform that had 20 ns rise/fall times (a thousand times faster than the 23 μs rise/fall times that characterize the 44.1 kHz sampling rate of the digital compact-disk).

The electronics used in this experiment was designed and built in-house because the required combination of response speed, linearity, power-supply stability, and damping ability (output impedance) was not found in commercial headphone-amplifiers. The result was an amplifier with input and output impedances of 1 MΩ and 50 mΩ respectively, a 3-dB power bandwidth of 0–2.2 MHz, a rise/fall time of 90 ns, and dc offset voltages under 0.6 mV at all stages. The linearity of the entire signal chain is essentially perfect (non-linear errors have sound levels of less than 0 dB SPL; please see below).

A functional diagram of the overall configuration is shown in Fig. 1. Internal to the amplifier, sandwiched between two buffers (B), is an RC low-pass filter and a switch S. The distance from the input sockets to the output phone socket is only 10 cm and the crucial signal path that includes the two buffers, filter, and switch is about 2.5 cm long. This compactness together with close-returned wiring eliminates inductive transients during switching. Teflon insulated point-to-point wiring minimized stray capacitances. Switch S selects/deselects the 10 nF polystyrene capacitor

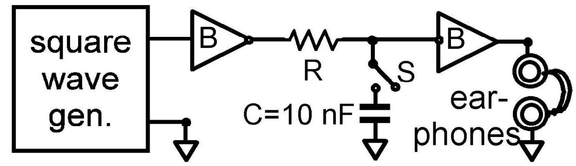

FIG. 1: *Functional diagram of circuitry.*

to include/exclude the low-pass filter (in the open position, the capacitance reduces to 0.1 nF). This capacitor together with the metal-film resistor R determines the τ=RC low-pass relaxation time constant, which is changed for each trial by replacing the resistor R with the requisite value. When the filter is switched off, the time constant is essentially zero ($\sim$0.1 μs). Altogether eight time constants were used in the trials (0, 3.9, 4.7, 5.6, 6.8, 7.7, 10, and 30 μs); the steps correspond to the available values of suitable metal-film resistors and adequately cover the time range of interest. The buffers have an input resistance of $R_{in} \approx 10^{13}$ Ω and an input capacitance of $C_{in} = 2$ pF. Thus there is negligible loading on the RC circuit and it performs in a nearly ideal manner.

The earphones used were a pair of Grado RS1 (Grado Laboratories, Brooklyn, New York) supra-aural headphones which have a frequency response of 12 Hz–30 kHz, an input resistance of 32 Ω, and an efficiency of 98 dB/mW. Identical signals are fed to both left and right ears to provide a diotic presentation.

B. Gating of signals

The method of gating used to transition between stimuli is a matter of crucial importance. A common practice for avoiding switching transients, is to gradually ramp down the first signal and then ramp up the second one. However, digital signal processing or voltage/computer controlled attenuators needed for signal ramping severely degrade the signal, especially in its temporal aspect. For the time scales and sensitivity sought in this work, these ramping devices would simply obliterate the small temporal differences being studied.

Special measures were taken in the design of the apparatus to minimize switching transients: relays (which can induce a disturbance through magnetic flux change) were avoided; stray inductance and capacitance, ground potentials, and DC offset voltages were prevented by design and optimized layout; switch-bounce distortion was avoided by minimizing contact areas and switching in the opening direction. These cumulative measures eliminated gating transients as can be seen from the waveforms of Fig. 2(a)–(c) (as a counterexample, panel (d) illustrates an example of poor quality switching where there is bounce). The transitions are almost mathematically perfect (harmonic amplitudes increase by the fraction $\sqrt{1+[2\pi f\tau]^2}$ and their phases shift by $\tan^{-1}[2\pi f\tau]$) and the unfiltered waveform continues without any anomaly, at the instant of switching, from where the filtered waveform terminated.

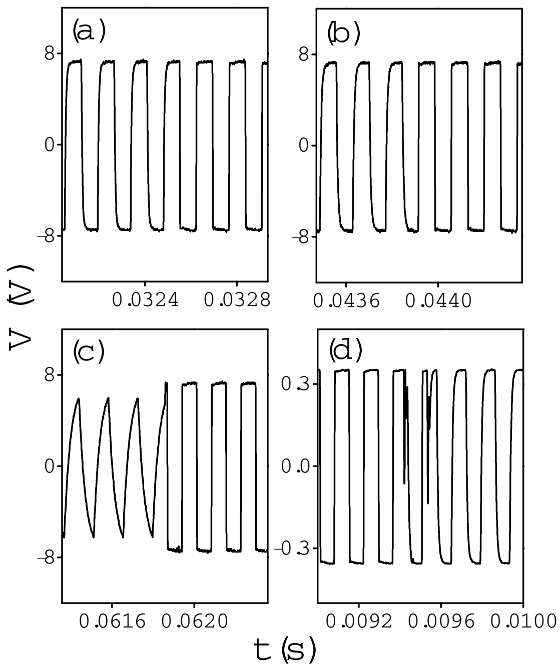

FIG. 2: *(a)–(c) Oscilloscope traces for clean switch openings (for $\tau = RC =$ 3.9, 5.6, and 30 μs respectively) showing perfect continuity of waveforms. (d) An oscilloscope trace in the case of ill-designed gating that shows waveform distortion due to switch-contact bounce.*

There is no ringing, overshoot, or other artifact, and the moving spectrum is perfectly continuous.

Besides the above electrical and waveform analyses, the inaudibility of gating cues was further investigated by two very effective blind listening tests, which are described in a later section.

C. Stimuli

The acoustic output from the earphone was measured with a flat-plate coupler using an ACO Pacific (ACO Pacific, Inc., Belmont, California) model 7016 measurement microphone and a model 4012 preamplifier with a 40 dB gain stage. The output of the preamplifier was fed to a LeCroy model LT322 (LeCroy Corporation, Chestnut Ridge, New York) 500 MHz digital storage oscilloscope, which digitized the signal at a sampling rate of 20 MS/s (million samples per second) and 12-bit vertical resolution. The frequency response of the entire measurement chain was flat (±3 dB) within 4 Hz–120 kHz. 240,000 traces were accumulated and averaged for each waveform to improve the signal-to-noise ratio.

Fig. 3 shows these measured waveforms. For the sake of clarity, the main panel shows only the waveforms for τ=0 (control), τ=4.7 (the discrimination threshold), and τ=10 μs (well above the threshold); the inset shows the entire set of waveforms for all τ. As expected, the waveform becomes progressively more rounded and attenuated as the filter time constant is increased. Note that there isn't any anomaly associated with the transient response of the earphone leading to discontinuously excessive ringing at

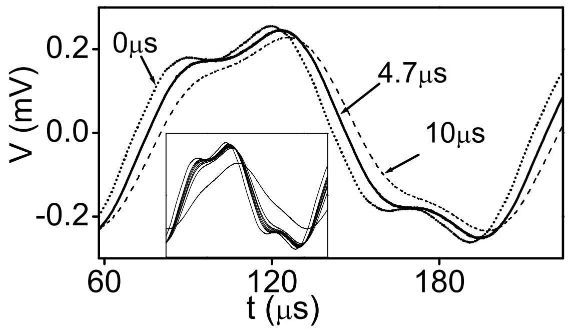

FIG. 3: *Waveforms of the acoustic output in the vicinity of the discrimination threshold ($\tau = 4.7\mu s$). The inset shows the complete set of waveforms (axes scales are identical) for all values of τ (0–30 μs from left to right). A Fourier analysis of the waveforms is given in Table I.*

lower time constants (which might aid in the discernment of the different stimuli). This suppression of ringing and well controlled transduction is an outcome of the amplifier's unusually low (50 mΩ) output impedance and consequent exceptional damping.

The presence of ultrasonic components at high transducer driving levels and inadequate damping can produce anharmonic distortion products within the audible band (Ashihara and Kiryu, 2000). To rule out the role of anharmonic distortion, subharmonics, noise, and other spurious components in the audible band, the spectrum of the signal was also measured separately using unaveraged signals (since synchronized averaging attenuates anharmonic frequencies). This power spectrum is shown in Fig. 4. No subharmonic peaks could be distinguished from noise; the absolute sound level of this noise in the 3.5 kHz subharmonic vicinity is <0 dB SPL (its fractional change due to filtering at threshold would be 0.046 dB). The collective power in all sub-fundamental frequencies in the 20 Hz–6 kHz band is less than 0.04% of the power in just the fundamental peak (7 kHz ± 0.1 kHz band). This freedom from anharmonic distortion is ensured by the superior quality of the electronics and transducer, and the relatively moderate (69 dB SPL) level of operation. Also because the entire signal chain is analog, spurious frequencies that can result from aliasing in digital systems are avoided. As will be shown below, the linearity—as reflected by the attenuations in the harmonic coefficients—is essentially perfect.

In view of the periodicity of the stimuli as established above, the waveforms (Fig. 3) can now be represented by a discrete Fourier series and are completely specified through the coefficients C_n and phases θ_n in the expansion $V(t) = \Sigma C_n \cos(2\pi f_n t + \theta_n)$, where $f_n = n \times 7$ kHz. These coefficients are given in Table I for all waveforms, and have been normalized with respect to the first harmonic of the control waveform of τ=0. The phase of each harmonic is specified with respect to its fundamental for the same τ value; the absolute phase and the phase difference across different values of τ are of course inconsequential. The noise floor of the coefficients is about 0.0003 and corresponds roughly to the quantization error of 1/4096 resulting from the 12-bit digital conversion. This error mag-

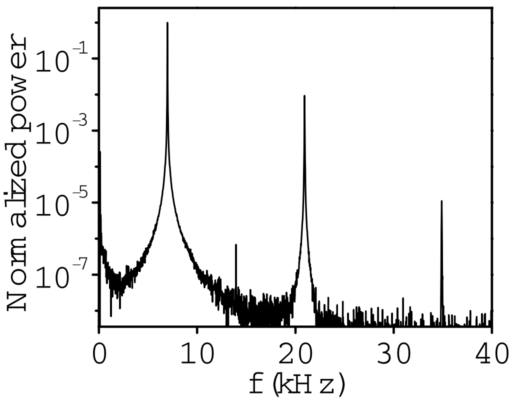

FIG. 4: *Power spectrum of the (unaveraged) acoustic output from an earphone (normalized by the power coefficient of the fundamental peak) in 20 Hz steps taken at a 2MS/s sampling rate.*

nitude corresponds to an absolute sound level of -2 dB SPL and is therefore negligible.

TABLE I: *Harmonic contents of acoustic signals. Coefficients $C_n(\tau)$ are expressed as a fraction of $C_1(\tau$=0). Phases θ_n, in radians, are expressed relative to the θ_1 for the same τ value.*

τ	f_1=7 kHz		f_3=21 kHz		f_5=35 kHz	
(μs)	C_1	θ_1	C_3	θ_3	C_5	θ_5
0	1.0000	0.00	0.2167	2.40	0.0183	1.89
3.9	0.9825	0.00	0.1921	2.36	0.0139	1.73
4.7	0.9791	0.00	0.1843	2.34	0.0128	1.64
5.6	0.9715	0.00	0.1753	2.31	0.0116	1.52
6.8	0.9580	0.00	0.1632	2.25	0.0102	1.35
7.7	0.9503	0.00	0.1583	2.21	0.0094	1.18
10	0.9147	0.00	0.1340	2.07	0.0076	0.84
30	0.6070	0.00	0.0564	0.96	0.0030	4.81

The acoustic output from the transducer is devoid of even numbered harmonics because of the square-wave signal fed to it. The total sound level at τ=0 is 69 dB SPL. The partial sound level of an nth harmonic at a particular value of τ can be obtained from the table using the formula: $L_p(f_n,\tau) \simeq 69 + 10\log[C_n^2(\tau)/\{\sum_{i=1}^{13} C_i^2(0)\}]$ dB; for example $L_p(14\text{kHz}, 0\mu\text{s}) \approx 0$ dB and $L_p(21\text{kHz}, 0\mu\text{s}) \simeq 55.5$ dB for τ=0, and $L_p(21\text{kHz}, 4.7\mu\text{s}) \simeq 54.1$ dB for τ=4.7 μs. Thus the levels of all harmonics beyond 7 kHz at all τ fall below their thresholds of audibility (International Standards Organization [ISO], 1996; Kurukata *et al.,* 2005; Ashihara *et al.,* 2006); the present subjects have measured high-frequency audibility limits of <18 kHz at 69 dB (details given below).

The measured attenuations in harmonic levels: $|\Delta L_p| = -20\log[C_n]$, conform exactly, within error bar, with the expected response for a first-order low-pass RC filter: $|\Delta L_p| = -10\log[1 + (2\pi f\tau)^2]$, attesting to the linearity of the electronics and transducer in this experiment. For example, the measured attenuations at 7 and 21 kHz for τ=4.7 μs are 0.183 dB and 1.41 dB, whereas the respective theoretical values are 0.182 dB and 1.41 dB.

Table II gives the measured attenuations in the fundamental and total rms values for different τ values. At the discrimination threshold of τ=4.7 μs, the drop in the first harmonic level is 0.18 dB (a 4.1% decrease in intensity) and the drop in the total rms value (across the whole spectrum) is ΔL_p= 0.23 dB (a 5.2% decrease in intensity). The just noticeable difference (JND) for the conditions in the experiment ($f \geq 7$ kHz and L_p=69 dB) is known (from Jesteadt, Wier, and Green, 1977) to be 0.7 dB (a 15% decrease in intensity). Even the 3 standard-error lower limit of this JND is 0.5 dB (an 11% decrease in intensity). Thus in the present experiment, differences in levels and spectral weights between the threshold and control stimuli seem too small for the discrimination to arise solely from direct spectral amplitude changes. As is discussed below, nonlinear mixing and an interference effect between quadratic and combination tones may contribute to the discernment.

TABLE II: *Signal attenuation. The first row gives the values of time constants. The second row gives the corresponding total (over entire spectrum) signal rms values as fractions of the value at τ=0. The third row gives the corresponding total attenuation in dB; the absolute sound level at τ=0 is 69 dB SPL. The fourth row gives the attenuation of the first harmonic (w.r.t. its value at τ=0); i.e., the entries correspond to $-20\log C_1$ for the C_1 values given in Table I.*

τ (μs)	0	3.9	4.7	5.6	6.8	7.7	10	30
rms value	1	0.978	0.974	0.965	0.950	0.942	0.903	0.596
rms att (dB)	0	0.19	0.23	0.31	0.45	0.52	0.88	4.50
7kHz att (dB)	0	0.15	0.18	0.25	0.37	0.44	0.77	4.34

It should be noted that the waveform at the eardrum will have a lower third-harmonic to first-harmonic ratio than the $C_3(\tau)/C_1(\tau)$ ratios of Table I, because of filtering by the ear canal. However, the fractional changes in each amplitude will be exactly the same as the $C_1(\tau)/C_1(0)$ and $C_3(\tau)/C_3(0)$ ratios of Table I. Thus the attenuations in the fundamental at the eardrum will be exactly the same as the values given in the fourth row of Table II, whereas the attenuation in the total rms signal will be marginally lower than the values in Table II's third row (the deficit not exceeding 0.05 dB at threshold). Thus the alteration in the signal's composition during its transit through the ear canal does not alter the conclusion drawn in the previous paragraph regarding the roles of acoustic spectral amplitudes. A detailed account of external- and middle-ear filtering is considered in a later section in which internal representations of stimuli are analyzed.

D. Procedure

1. *Main experiment*

In the main experiment, subjects try to discern differences between a (7 kHz approximately square-wave shaped) signal with finite low-pass filtering versus a control signal with no filtering (waveforms depicted in Fig. 3). The control tone was perceived to have a sharper or brighter timbre whereas the filtered one had a duller quality (no difference in loudness was perceived except for the largest setting of

τ=30 μs). In the blind test, the subject tries to judge whether an unknown sound is the control or filtered tone for different settings of τ. It was found in preliminary testing (especially when τ is close to the threshold) that subjects needed to listen to the tones for several seconds to form a lasting impression of the sounds; immediately after switching the subjects had difficulty assessing whether anything had changed or not. This again confirms that the gating itself does not provide a cue.

The time course of the trials is as follows. The appropriate resistor R is soldered into the low-pass filter stage (Fig. 1) to provide the required τ=RC time constant. The position of the switch S determines whether the tone being played is the control or filtered one (for this particular τ). The resistor and τ value are not changed until all five subjects have been tested for this τ. Then the electronics is dismantled and a new resistor is soldered in to change τ to the next value to be tested (using sockets or clips, in place of soldering, was found to degrade the circuit performance). For each τ, a subject listens to the control and filtered sounds several (3–10) times to become familiar with each sound. The sounds are now played in the sequence: filtered, unknown, and control. The duration of the filtered tone at the start of the sequence was limited to 20 s, and the durations of the unknown and control were each limited to 10 s. The gating between the signals was described earlier. The subject judges the identity of the unknown by comparing it to his or her recent memories of the known control and filtered sounds, after being allowed to listen to the entire sequence twice. Once the judgement has been recorded, the next trial for the same τ setting is conducted. For each trial, the unknown sound is chosen to be either filtered or control (with, on average, equal likelihood for each) depending on a random-number sequence generated by a computer. One example of such a sequence is $\{1, 0, 0, 1, 1, 1, 1, 0, 0, 1\}$. When all ten trials for one subject have been completed, a ten-trial set is conducted on the next subject (for this same τ setting). When all five subjects have been tested at this τ setting (completing a total of 50 blind trials), the electronics is modified and a new resistor corresponding to the next time constant is soldered into the circuit (the capacitor is kept the same for all settings). Now ten trials are conducted on each of the five subjects at this new τ setting, and so on. As τ is made progressively smaller, the distinction between control and filtered sounds is lost when the threshold is crossed. As mentioned earlier, the tests were carried out for seven finite values of the time constant: τ = 30, 10, 7.7, 6.8, 5.6, 4.7, and 3.9 μs, in addition to the control value of τ=0. Altogether the main experiment consists of 350 blind trials (50 for each τ setting). 50 additional blind trials were done on a second group of two new subjects at just τ=5.6 μs, to reinforce the statistics for this setting.

2. *Gating checks*

Two supplementary experiments were performed for the purpose of verifying the inaudibility of switching transients: (1) In the first, a 100 KΩ high quality noise-free potentiometer (by Noble U.S.A., Inc., Rolling Meadows, Illinois) was connected in series with the switch S (Fig. 1). In this variation of the main experiment, the switch remained in one position and the listeners task was to discern the effect of filtering when the potentiometer was turned from one side to the other (the transition was spread out gradually over $\sim$20,000 cycles).
(2) In the second supplementary experiment, the 7 kHz square-wave signal was replaced by an ultrasonic (22 kHz) one of the same level for a duration of 40 s. In the middle of this (after 20 s) the low-pass filter is switched out (or not), depending on a random number sequence, and the subject listened for any audible cue such as a transient. The ultrasonic signal allows switching transients to be easily heard (in the case of poor quality switching) because masking from the tone itself is absent.

E. Listeners

Seven listeners (Groups 1 and 2) participated in the main experiment, three (Group 3) in the gating-check using the alternative gradual-transition procedure, and five (Group 4) in the blind tests for the 22 kHz gating check. The ranges of ages of the listeners for the four groups were respectively 26–46, 19–27, 24–47, and 17–36 years, and none had a history of hearing impairment or neurological disease. Their high-frequency threshold of hearing, f_{max}, was measured using the same apparatus by slowly sweeping a pure sine-wave signal (instead of the square-wave signal) from 25 kHz to 5 KHz in 2 min while maintaining a fixed level of 69 dB SPL. These f_{max} values are shown in the first column of Table III. For each subject the threshold was measured 6 times resulting in a standard deviation of $\sigma\approx$0.2 kHz. As expected, no subject was able to hear upto 21 kHz even at L_p=69 dB (in the experiment, the level of this harmonic is $<$ 56 dB under all conditions).

The subjects in Groups 1, 3, and 4 were volunteers and were not paid. The two individuals in Group 2 were paid subjects. The University of South Carolina Institutional Review Board (IRB) reviewed and approved the proposal for this research activity and the requisite consent forms.

III. RESULTS

A. Experiment 1: Groups 1 and 2

Table III shows the results of the main experiment, performed on Group 1. All five subjects scored 100% on their blind tests for $\tau\geq$5.6 μs. The shortest time constant that could be readily discerned was τ= 4.7 μs. For this, combining all subjects, there were 86% correct judgements, a chi-squared analysis value of χ^2 = 25.9 (which well exceeds the critical value of 3.84 for one degree of freedom) and a discriminability index of d' = 2.26 with a criterion of c = 0.92. At 3.9 μs, essentially no difference could be heard (54% correct judgements, χ^2 = 0.32, d' = 0.23, and c = 0). Two additional subjects (Group 2) were tested at only 5.6 μs. Each one was given 25 blind trials, on which they both scored 100%.

TABLE III: *Results of blind trials of main experiment (Group 1). Each row corresponds to a different subject. The first column gives the high-frequency threshold for pure tones at the sound level of 69 dB SPL. Other columns correspond to a different τ as indicated at the top. The entries indicate the number of correct judgements (out of 10) for that τ.*

f_{max} (kHz)	**30** μs	**10** μs	**7.7** μs	**6.8** μs	**5.6** μs	**4.7** μs	**3.9** μs
17.8	10	10	10	10	10	10	5
16.6	10	10	10	10	10	10	5
17.7	10	10	10	10	10	8	8
14.8	10	10	10	10	10	8	5
9.4	10	10	10	10	10	7	4

B. Experiment 2: Group 3

Three additional subjects (Group 3) were tested using the slightly modified apparatus and alternative protocol whereby a potentiometer was used to vary the signal continuously from filtered to unfiltered, instead of switching. 40 blind trials were conducted on each of the three subjects, at τ=5.6 μs, resulting in 100% correct judgements. This not only proves the irrelevance of gating cues, but further reinforces the statistics at 5.6 μs.

C. Experiment 3: Group 4

A third experiment was done on Group 4 (5 subjects), to see whether they could hear a gating transient during a 22 kHz ultrasonic signal. Such a signal is itself inaudible allowing a switching transient, if any, to be clearly audible. However, the subjects could not demonstrably discern transients and perceive the switching event: 100 blind trials (20 per subject) were carried out resulting in an overall score of 46% correct judgements, a hit rate (=hits/[hits+misses]) of 0.06, false-alarm rate (=false alarms/[false alarms + correct rejections]) of 0.10, $d' = -0.32$, $c = 1.26$, and $\chi^2 = 0.64$.

IV. DISCUSSION AND CONCLUSIONS

A. Psychoacoustics and physiology

The present work investigated the temporal resolution of the human auditory system through the threshold time constant τ of low-pass filtering. The results on Group 1 indicate a threshold of around 4.7 μs. For τ=5.6 μs, combining the results on Groups 1–3, the cumulative score is 100% for a total of 220 blind trials carried out over 10 different individual subjects. This corresponds to χ^2=220 for audibility at τ=5.6 μs.

A key reason for the shorter threshold observed in the present work may be because the response of the electronics and overall signal chain is much faster and more precise than equipment typically used in previous psychoacoustic research. Notwithstanding, an even shorter threshold might be obtained using an even faster instrument chain (such as a higher bandwidth transducer) or younger and better trained subjects (the listeners in this study were faculty members and students from the Physics and Astronomy department and were not professionally involved with music or acoustics). Therefore the threshold value obtained here should be viewed as an upperbound rather than the final word on the temporal resolution. Nevertheless, this experiment clearly shows an effect at time scales about three times shorter than found in past investigations of temporal resolution that attempted to avoid spectral cues (Krumbholz et al., 2003, obtained a threshold of 12.5 μs). Besides the quantitative reduction of the upperbound, there is also the qualitative significance that the present threshold is, for the first time, less than the nominally expected $1/2\pi f_{max} \approx 9$ μs. This means that for some psychoacoustic experiments, test-equipment bandwidths may need to be higher than the maximum audible frequency to avoid inadvertent audible alteration of signals.

In the present experiment, one effect of low-pass filtering is to diminish all amplitudes by an extent that can be straightforwardly calculated as well as measured. The measurement shows that the decreases in levels (the total rms value of the signal as well as the change in any of the individual harmonics) lie below their JNDs so that the straighforward attenuations in a simple linear scenario should not be enough to determine the discrimination. On the other hand the filtering introduces significant frequency-dependent phase shifts (for τ=4.7 μs, the 21 kHz component shifts by 32°), spreading out the signal over a time scale τ and making the waveform shape more rounded as seen in Fig. 3.

How is this waveform change detected? Time-domain models (e.g., Meddis and Hewitt, 1991; Patterson *et al.*, 1992; Patterson, 1994; Patterson *et al.*, 1995; Meddis and O'Mard, 1997; Irino and Patterson, 1997; Krumbholz and Wiegrebe, 1998; Wiegrebe and Krumbholz, 1999; Irino and Patterson, 2001; Irino and Patterson, 2006) seek to trace the evolution of the signal as it progresses from the initial acoustic stage through the basilar-membrane motion (BMM) and hair-cell transduction to an internal neural representation. The outer- and middle-ear transfer functions are modeled as broadband filters: as second-order butterworth filters with cutoff frequencies of 450 and 8000 Hz (Meddis and O'Mard, 1997; Wiegrebe and Krumbholz, 1999); or as an inversion of the equal level contours (ELC), minimal audible field (MAP), or minimal audible pressure (MAP) curves (Glasberg and Moore, 1990). This broadband filtered sound then reaches the basilar membrane, whose tonotopy can be modeled as a bank of bandpass filters. The number of filters, their bandwidth, and their response functions differ between different specific models, such as the gammatone filter (Boer, 1975; de Boer and de Jongh, 1978; Patterson et al., 1992), dynamic-compressive gammachirp filter (Irino and Patterson, 2006), etc. The model should also allow for the generation of combination tones such $f_1 - n(f_2 - f_1)$ that can arise from non-linearities in the cochlear mechanics and then propagate to their appropriate frequency channel (Patterson et al., 1995). The BMM is transduced in the inner hair cells (IHC) into a receptor potential. These IHCs have a limited temporal speed as evidenced by loss of phase locking around 3–4 kHz (Johnson, 1974; Shamma, 1989) which leads to a smoothening of temporal fine structure. This effect can be incorporated in a model by low-pass filtering the output of the filters (Carney, 1993). Early work (e.g., Viemeister, 1979) assumed the lowpass cutoff to be 60 Hz whereas Palmer and Russel (1986) showed that cutoffs in the 600–2000 Hz range are more realistic; Krumbholz and Wiegrebe (1998) used a second order lowpass cutoff of 1.1 kHz in their analysis. The hair cells are contacted by auditory nerve fibers whereby the continuously variable IHC receptor potential is converted into stochastic nerve impulses (spikes). The nerve firing only takes place on positive half cycles of the IHC potential, which aspect is modeled as a half-wave rectification step. The next step of a model is to compute the spike probability for each filter channel as a function of that filter's output intensity. This conversion from the BMM to a neural activity pattern (NAP) may be based on IHC simulation (Meddis, 1988) or a functional (e.g., adaptive thresholding) mechanism (Patterson et al., 1995). IHC/auditory-nerve adaptation that results from depletion of neurotransmitter is taken into account in models such as Sumner (2002). In addition, the auditory periphery undergoes adaptation on longer time scales, which has been incorporated by feedback loops with time constants in the 5–500 ms range (Kohlrausch and Püschel, 1988; Kohlrausch et al., 1992; Dau et al., 1996). Once the NAP has been computed for two stimuli, the corresponding discriminability index can be estimated from the correlations and variations between different NAP instances for each stimulus and a template (average over several instances of the control pattern). Variability between different NAP instances for the same stimulus arises from the the spontaneous discharge rate and stochasticity of nerve firing. This randomness can be incorporated into a model through Gaussian noise (Dau et al., 1996). For certain temporal tasks (e.g., edge detection) it may be appropriate to sum over different channels of an NAP; it is known from physiology that such synchronous cross-frequency comparisons take place in the octopus neurons in the posteroventral cochlear nucleus (Golding et al., 1995; Ferragamo and Oertel, 1998; Golding et al., 1999; Oertel et al., 2000).

Some examples of recent experiments that probed temporal integration (temporal window Δt over which signal energy is summed up) and temporal resolution (the ability to distinguish quick fluctuations in the instantaneous signal amplitude), in which the results could be well fitted by models, are the works by Krumbholz and Wiegrebe (1998) and Wiegrebe and Krumbholz (1999) respectively. The first work measured the lowering of threshold for detecting two tone or noise bursts as they are brought close together in time. The authors showed that the interaction between the brief signals was the result of temporal overlap of auditory filter responses prior to mechanical-to-neural transduction (based on the observations that the threshold temporal separation scaled inversely with frequency and depended on the relative phases). The time scale for this overlap varies inversely with frequency, as mentioned above, and at low frequencies was as large as 10–20 ms. In their other work on temporal resolution (Wiegrebe and Krumbholz, 1999) listeners tried to distinguish various equal-energy combinations of noise pips (e.g., a single pip with a steady state portion versus two pips with equal peak to peak separation). The shortest threshold observed was ~0.5 ms. One of the conclusions of this work was again that the main limitation imposed on the perception of transient stimuli occurs in the peripheral auditory system; it was expected that central processing limitations should take over only for long duration stimuli. They found that their obervations could be fit by modeling the peripheral auditory filtering with a gammatone filterbank.

In the present experiment, the two acoustic stimuli being compared are both long-duration steady complex tones whose essential compositions are well approximated by (from Table I): $K''[0.98\cos(2\pi 7000t)+0.18\cos(2\pi 21000t+\phi''_A)]$ for Tone A (4.7 μs filtered) and $K''[\cos(2\pi 7000t)+0.22\cos(2\pi 21000t+\phi''_B)]$ for Tone B (unfiltered). By the time the signals arrive at the cochlea, external- and middle-ear filtering change these signals to (applying an "ELC" correction as per Glasberg and Moore, 1990): $K'[0.98\cos(2\pi 7000t)+0.15\cos(2\pi 21000t+\phi'_A)]$ for Tone A (4.7 μs filtered) and $K'[\cos(2\pi 7000t)+0.19\cos(2\pi 21000t+\phi'_B)]$ for Tone B (unfiltered). The weak 21 kHz component is far outside the bandwidth of the highest filterbank channel and so we will assume that it will not directly contribute to the NAP. However nonlinearities in cochlear mechanics and the preceding mechanical chain can generate an audible 14 kHz component. For a nonlinear response represented by $y \propto x + bx^2$ (with $b < 0$ for a compressive nonlinearity and $b^2 \sim 1/100$ as per Zwicker, 1981), an input consisting of a fundamental and third harmonic mixture $x \propto \cos\omega_0 t + a\cos(3\omega_0 t+\theta)$ will give rise to a response $y \propto \cos\omega_0 t + \frac{b}{2}\cos(2\omega_0 t) + ab\cos(2\omega_0 t+\theta)$ (keeping oscillating terms up to $2\omega_0$ in frequency). The second term, with $2\omega_0$, comes from doubling the fundamental and maintains the same phase; the last term with $2\omega_0$ arises as a difference tone between ω_0 and $3\omega_0$ (in the input) and maintains their original phase difference θ. Applying this nonlinearity to the previous middle-ear filtered signals gives the effective signals feeding the BMM filterbanks: $K[0.98\cos(2\pi 7000t) + b\{0.5\times 0.98^2\cos(2\pi 14000t) + 0.15\times 0.98\cos(2\pi 14000t+\phi_A)\}]$ (Tone A) and $K[\cos(2\pi 7000t) + b\{0.5\cos(2\pi 14000t)+0.19\cos(2\pi 14000t+\phi_B)\}]$ (Tone B). The phase difference $\Delta\phi = \phi_B - \phi_A = \phi''_B - \phi''_A = \{\tan^{-1}(-2\pi 7000\tau) - \tan^{-1}(-2\pi 21000\tau)\} = 20^{\circ}$ is preserved during the nonlinear mixing as explained earlier. The maximum difference in levels of the nonlinearly generated second harmonic between the tones A and B can now be calculated: $\Delta L_p(14\text{ kHz}) = 10\log([0.5+0.19\sin(\Delta\phi/2)]^2/[0.5\times 0.98^2+0.15\times 0.98\sin(-\Delta\phi/2)]^2) = 1.4$ dB. This level change is eight times larger than the subliminal $\Delta L_p(7\text{ kHz}) = 0.18$ dB of the fundamental, thus allowing a listener to discern the phase shift of the 21 kHz component (and hence waveform shape) indirectly through an interference between the two (quadratic and difference) 14 kHz nonlinear products. It should be noted that while this model is offered as one possible explanation (which can be valid for listeners whose upper audibility threshold is over 14 kHz) there may be other explanations as well and it is hoped that these experimental results will provide stimulation for auditory theorists.

There is an important distinction between the present experiment, in which the phase of a very modest level of ultrasound (21 kHz at 55 dB) is possibly detected by a novel interference effect between two distinct (quadratic and difference) coherent audible (14 kHz) nonlinear products, and experiments (Henry and Fast, 1984; Ashihara *et al.,* 2006) that demonstrate audibility of very high level (> 85 dB) ultrasound by itself, presumably through the generation of audible subharmonics (von Gierke, 1950; Ashihara *et al.,* 2006). In the latter case, the coefficient for subharmonically generated power $\sim 10^{-8}$ (from their observed threshold of $\sim$85 dB for 21 kHz) is many orders of magnitude lower than the nonlinear power coefficient $b^2 \sim 1/100$ in the present experiment. Thus while an ultrasonic frequency at a modest level might not be audible by itself, the ear seems to be sensitive to alterations in its strength as well as phase when it is present as a harmonic of an audible fundamental. For such stimulus alterations to be heard, it is necessary for the instrumentation to be fast enough to resolve the difference and preserve the phase information. It is hoped that this result will bring awareness of the possible benefits of higher instrumentation bandwidths for certain psychoacoustic experiments and that it will encourage researchers to further inquire into the auditory system's nonlinear behavior and its response to temporal convolution.

B. Implications for sound reproduction

The result presented here has relevance for the performance requirements of audio components and digital encoding schemes. It is known that the bandwidth requirement for sonically transparent audio reproduction is higher than the 20 kHz: in the coding of digital audio it has been noted (Stuart, 2004) that listeners show a preference for a 96 kHz sampling rate over the CD (digital compact disk) standard of 44.1 (i.e., a 22 kHz Nyquist frequency). It is sometimes thought that this may be due to the less drastically sloped cutoff of the digital filter and the reduced disturbances introduced in the audible pass band. The present work shows that the bandwidth requirement into the ultrasonic range is more fundamental and not just due to artifacts of digital filtering. It is also commonly conjectured in the audio literature that the time-domain response of a system (e.g., temporal smearing caused by capacitive and other energy-storage mechanisms in cables) is a key factor in determining the transparency of reproduction (see, for example, van Maanen, 1993). However a search of the literature revealed an absence of a controlled blind experiment comparable to the one conducted here. The present work thus contributes toward a better fundamental understanding and provides a quantitative measure for audio-reproduction standards.

V. POST SCRIPT ADDED AT PROOF STAGE

In a sister experiment (Kunchur, 2007), where signals were temporally altered by spatial displacement of speakers instead of by electronic means, a similar threshold of 6 μs was found. That work also makes a rudimentary neurophysiological estimate for the temporal resolution for transient stimuli that is in the 2–16 μs range.

VI. ACKNOWLEDGEMENTS

The author gratefully acknowledges G. Saracila and M. W. Rhoades for help with literature searching and other assistance, and acknowledges useful discussions and feedback from J. M. Knight, W. M. Hartmann, R. Meddis, F. G. Zeng, D. Oertel, E. W. Healy, R. A. Lutfi, R. D. Patterson, B. C. J. Moore, W. A. Yost, A. Botero, D. H. Wedell, K. Kubodera, R. White, and D. H. Arcos. This work was partially supported by a grant from the University of South Carolina Office of Research and Health Sciences Research Funding Program.

[1] Ashihara, K., and Kiryu, S. (2000). "Influence of expanded frequency band of signals on non-linear characteristics of loudspeakers", Nippon Onkyo Gakkai Shi (J. Acoust. Soc. Jap.) **56**, 549–555.

[2] Ashihara, K., Kurukata, K., Mizunami, T., and Matsushita, K. (2006). "Hearing threshold for pure tones above 20 kHz", Acoust. Sci. & Tech. **27**, 12–19.

[3] Carney, L. H. (1993). "A model for the response of low-frequency auditory-nerve fibers in cat", J. Acoust. Soc. Am. **93**, 401–417.

[4] Corso, F. J. (1963). "Bone conduction thresholds for sonic and ultrasonic frequencies", J. Acoust. Soc. Am. **35**, 1738–1743.

[5] Dau, T., Püschel, D., and Kohlrausch, A. (1996). "A quantitative model of the 'effective' signal processing in the auditory system: II. Simulations and measurements", J. Acoust. Soc. Am. **99**, 3623–3631.

[6] Deatherage, B. H., Jeffress, L. A., and Blodgett, H. C. (1954). "A note on the audibility of intense ultrasound", J. Acoust. Soc. Am. **26**, 582.

[7] de Boer, E. (1975). "Synthetic whole-nerve action potentials for the cat", J. Acoust. Soc. Am. **58**, 1030–1045.

[8] de Boer, E. and de Jongh, H. R. (1978). "On cochlear encoding: Potentialities and limitations of the reverse-correlation techniques", J. Acoust. Soc. Am. **63**, 115–135.

[9] Eddins, D. A., Hall, J. W., and Grose, J. H. (1992). "Detection of temporal gaps as a function of frequency region and absolute bandwidth", J. Acoust. Soc. Am. **91**, 1069–1077.

[10] Ferragamo, M. J., and Oertel, D. (1998). "Shaping of synaptic responses and action potentials in octopus cells", J. Assoc. Res. Otolaryngol. **21**, 96.

[11] Glasberg, B. R., and Moore, B. C. J. (1990). "Derivation of auditory filter shapes from notched-noise data", Hearing Res. **47**, 103–138.

[12] Gockel, H., Moore, B. C. J., Plack, C. J., Carlyon, R. P. (2006). "Effect of noise on the detectability and fundamental frequency discrimination of complex tones", J. Acoust. Soc. Am. **120**, 957–965.

[13] Golding, N. L., Robertson, D., Oertel, D. (1995). "Recordings from slices indicate that octopus cells of the cochlear nucleus detect coincident firing of auditory nerve fibers with temporal precision", J. Neurosci. **15**, 3138–3153.

[14] Golding, N. L., Ferragamo, M., and Oertel, D. (1999). "Role of intrinisic conductances underlying transient response of

octopus cells of the cochlear nucleus", J. Neurosci. **19**, 2897–2905.

15 Hall, D. E. (2002). "Chapter 6: The human ear and its response" in *Musical Acoustics*, Third Edition, Brooks/Cole Thompson Learning Publishing, pp. 94.

16 Henning, B. G. (1974). "Detectability of interaural delay in high-frequency complex waveforms", J. Acoust. Soc. Am. **55**, 84–90.

17 Henning, G. B., and Gaskell, H. (1981). "Monaural phase sensitivity with Ronken's paradigm", J. Acoust. Soc. Am. **70**, 1669–1673.

18 Henry, K. R., and Fast, G. A. (1984). "Ultrahigh-frequency auditory thresholds in young adults: Reliable responses up to 24 kHz with a quasi-free field technique", Audiology **23**, 477–489.

19 Irino, T., and Patterson, R. D. (1997). "A time-domain, level-dependent auditory filter: The gammachirp", J. Acoust. Soc. Am. **101**, 412–419.

20 Irino, T., and Patterson, R. D. (2001). "A compressive gammachirp auditory filter for both physiological and psychophysical data", J. Acoust. Soc. Am. **109**, 2008–2022.

21 Irino, T., and Patterson, R. D. (2006). "A dynamic compressive gammachirp auditory filterbank", IEEE Trans. on Audio, Speech, and Language Processing **14**, 2222–2232.

22 International Standards Organization minimum audible field (MAF) standard: ISO 389-7 (1996).

23 Jesteadt, W., Wier, C. C., and Green, D. M. (1977). "Intensity discrimination as a function of frequency and sensation level", J. Acoust. Soc. Am. **61**, 169.

24 Johnson, D. H. (1974). "The response of single auditory-nerve fibers in the cat to signle tones: synchrony and average discharge rate", Ph.D. thesis, Department of Electrical Engineering, MIT, Cambridge, MA.

25 Kohlrausch, A., and Püschel, D. (1988). "Interrelations between a psychoacoustical model of temporal effects in hearing and neurophysiological observations", in *Sense Organs*, edited by N. Elsner and F. G. Barth (Thieme, Stuttgart), p. 39.

26 Kohlrausch, A., Püschel, D., and Alphei, H. (1992). "Temporal resolution and modulation analysis in models of the auditory system", in *The Auditory Processing of Speech*, edited by M. E. H. Schouten (Mouton de Gruyter, Berlin, New York), pp. 85–98. (1988).

27 Krumbholz, K., and Wiegrebe, L. (1998). " Detection threshold for brief sound — are they a measure of auditory intensity integration?", Hearing Res. **124**, 155–169.

28 Krumbholz, K., Patterson, R. D., Nobbe, A., and Fastl, H. (2003). "Microsecond temporal resolution in monaural hearing without spectral cues?", J. Acoust. Soc. Am. **113**, 2790–2800.

29 Kunchur, M. N. (2007). "Audibility of temporal smearing and time misalignment of acoustic signals", Electronic Journal Technical Acoustics, http://www.ejta.org, 2007, 17. (Preprint can be downloaded from http://www.physics.sc.edu/kunchur/align.pdf)

30 Kurukata, K., Mizunami, T., Matsushita, K., and Ashihara, K. (2005). "Statistical distribution of normal hearing thresholds under free-field listening conditions", Acoust. Sci. & Tech. **26**, 440–446.

31 Lenhardt, M. L., Skellett, R., Wang, P., and Clarke, A. M. (1991). "Human ultrasonic speech perception", Science **253**, 82–85.

32 Lenhardt, M. L. (1998). "Human ultrasonic hearing", Hearing Rev. **5**, 50–52.

33 B. Leshowitz, "Measurement of the two-click threshold", J. Acoust. Soc. Am. **49**, 462–466 (1971).

34 van Maanen, H. R. E. (1993). "Temporal decay: a useful tool for the characterisation of resolution of audio systems?", AES Preprint 3480 (C1-9), presented at the 94th convention of the Audio Engineering Society in Berlin.

35 Meddis, R. (1988). "Simulation of auditory-neural transduction: Further studies", J. Acoust. Soc. Am. **83**, 1056-1063.

36 Meddis, R., and Hewitt, M. J. (1991). "Virtual Pitch and phase sensitivity of a computer model of the auditory periphery. I. Pitch identification", J. Acoust. Soc. Am. **89**, 2866–2882.

37 Meddis, R., and O'Mard, L. (1997). " A unitary model of pitch perception", J. Acoust. Soc. Am. **102**, 1811–1820.

38 Moore, B. C. J., Glasberg, B. R., Flanagan, H. J., and Adams, J. (2006). "Frequency discrimination of complex tones; assessing the role of component resolvability and temporal fine structure", J. Acoust. Soc. Am. **119**, 480–490.

39 J. O. Nordmark, "Binaural time discrimination", J. Acoust. Soc. Am. **35**, 870–880 (1976).

40 Oertel, D., Bal, R., Gardner, S. M., Smith, P. H., and Joris, P. X. (2000). "Detection of synchrony in the activity of auditory nerve fibers by octous cells of the mammalian cochlear nucleus", Proc. Nat. Acad. Sci. **97**, 11773–11779.

41 Oohashi, T., Nishina, E., Kawai, N., Fuwamoto, Y., and Imai, H. (1991). "High-frequency sound above the audible range affects brain electric activity and sound perception", J. Audio Eng. Soc. (Abstracts) **39**, 1010.

42 Palmer, A. R. and Russell, I. J. (1986). "Phase locking in the cochlear nerve of the guinea pig and its relation to the receptor potential of inner hair cells", Hearing Res. **24**, 1–15.

43 Patterson, R. D., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C., and Allerhand, M. (1992). "Complex sounds and auditory images", in *Auditory Physiology and Perception, Proceedings for the 9th International Symposium on Hearing*, edited by Y. Cazals, L. Demany, and K. Horner (Pergamon, Oxford), pp. 429–446.

44 Patterson, R. D. (1994). "The sound of a sinusoid: Time-interval models", J. Acoust. Soc. Am. **96**, 1419–1428.

45 Patterson, R. D., Allerhand, M., and Giguere, C. (1995). "Time domain modeling of peripheral auditory processing: A modular architecture and software platform", J. Acoust. Soc. Am. **98**, 1890-1894.

46 Patterson, R. D., and Datta, A. J. (1996). "The detection of iterated rippled niose (IRN) masked by IRN", Br. J. Audiol., **30**, 148.

47 Penner, M. J. (1977). "Detection of temporal gaps in noise as a measure of the decay of auditory sensation", J. Acoust. Soc. Am. **61**, 552–557.

48 Plomp, R. (1964). "Rate of decay of auditory sensation", J. Acoust. Soc. Am. **36**, 277–282.

49 Pumphrey, R. J. (1950). "Upper limit of frequency for human hearing", Nature **166**, 571.

50 Ronken, D. (1970). "Monaural detection of a phase difference between clicks", J. Acoust. Soc. Am. **70**, 1091–1099.

51 Shamma, S. A., Shen, N., Preetham, G. (1989). "Stereausis: Binaural processing wihtout neural delays", J. Acoust. Soc. Am. **86**, 989–1006.

52 Stuart, J. R. (2004). "Coding for high-resolution audio systems", J. Audio Eng. Soc., **52**, 117–144.

53 Sumner, C. J., Lopez-Poveda, E. A., O'Mard, L. P., Meddis, R. (2002). "A revised model of the inner-hair cell and auditory-nerve complex", J. Acoust. Soc. Am. **111**, 2178–2188.

54 Viemeister, N. F. (1979). "Temporal modulation tranfer functions based upon modulation thresholds", J. Acoust. Soc. Am. **66**, 1364–1380.

55 von Gierke, H. E. (1950). "Subharmonics generated in human and animal ears by intense sound", J. Acoust. Soc. Am. **22**, 675.

[56] Wiegrebe, L., and Krumbholz, K. (1999). "Temporal resolution and temporal masking properties of transient stimuli: Data and an auditory model", J. Acoust. Soc. Am. **105**, 2746–2756.

[57] Yoshikawa, S., Noge, S., Ohsu, M., Toyama, S., Yanagawa, H., Yamamoto, T. (1995). "Sound-quality evaluation of 96-kHz sampling digital audio", J. Audio Eng. Soc. (Abstracts) **43**, 1095.

[58] Yost, W. A., Patterson, R. D., and Sheft, S. (1996). "A time domain description for the pitch strength of iterated rippled noise", J. Acoust. Soc. Am. **99**, 1066–1078.

[59] Zwicker, E. (1981). "Formulae for calculating the psychoacoustical excitation level of aural difference tones measured by the cancellation method", J. Acoust. Soc. Am. **69**, 1410–1413.