

Audibility of temporal smearing and time misalignment of acoustic signals

Milind N. Kunchur*
Department of Physics and Astronomy
University of South Carolina, Columbia, SC 29208
(Dated: 19.07.2007 [received]; 29.08.2007 [published])

Misalignment in timing between drivers in a speaker system and temporal smearing of signals in components and cables have long been alleged to cause degradation of fidelity in audio reproduction. It has also been noted that listeners prefer higher sampling rates (e.g., 96 kHz) than the 44.1 kHz of the digital compact disk, even though the 22 kHz Nyquist frequency of the latter already exceeds the nominal single-tone high-frequency hearing limit f_{max}∼18 kHz. These qualitative and anecdotal observations point to the possibility that human hearing may be sensitive to temporal errors, τ, that are shorter than the reciprocal of the limiting angular frequency $[2\pi f_{max}]^{-1} \approx 9$ μs, thus necessitating bandwidths in audio equipment that are much higher than f_{max} in order to preserve fidelity. The blind trials of the present work provide quantitative proof of this by assessing the discernability of time misalignment between signals from spatially displaced speakers. The experiment found a displacement threshold of d≈2 mm corresponding to a delay discrimination of τ≈6 μs.

Keywords: Time, temporal, align, alignment, smearing, resolution.

INTRODUCTION

An ideal sound-reproduction chain will reproduce an exact replica of the original acoustic signal received by the recording microphone. In this case, the final reproduced acoustic signal will have the same waveform shape in the time domain as well as an identical Fourier spectrum. For practical reasons a reproduction chain cannot have an infinitely fast response time nor an infinitely wide frequency bandwidth; however, these idealizations are neither necessary nor desirable since the ear has its own limitations in both of these domains.

Usually the response time τ and high-frequency bandwidth limit f_{max} go hand-in-hand. For linear systems, there is a reciprocal relationship between τ and ω_{max} $(=2\pi f_{max})$, the angular-frequency limit. Thus an input signal represented by a narrow pulse will produce an output signal spread out over a characteristic time of $\tau \sim 1/\omega_{max}$. However, the mechanism of hearing is complicated and non-linear. Furthermore, the extraction of spectral and temporal information involves entirely separate chains of neural circuitry with different neuron types (having different response speeds) and different neural-circuit topologies. As a result the temporal resolution is not directly related to the highest audible frequency and the limiting τ can be much shorter than $1/\omega_{max}$.

In a sound-reproduction system also, complexities in the response (such as due to dielectric relaxation, mechanical vibrations in cables, reverberation within speaker cabinets, and other mechanisms that store and slowly release energy) invalidate its categorization as a perfect linear system, which in turn negates a simple connection between τ and $1/\omega_{max}$. Because of this, an adequate frequency response need not ensure that a component will be sonically transparent (i.e., not produce an audible degradation of the signal). It is recognized in the audio community that smearing in the time-domain is a key factor in degrading transparency [1, 2] and that temporal misalignment can produce audible errors in the spectral response [3]. Besides such scientific works, popular literature and advertisements targeted toward high-end audio enthusiasts abound with claims of the importance of "time alignment" of speaker drivers (for which reason some models of speakers have slanted front baffles), the "coherence" of small speakers over large speakers (hence the preference of some listeners for compact monitors over large dipole panels), "time coherence" in cables (i.e., avoiding dispersion so that all frequencies arrive together), and a myriad other effects related to smearing in the time domain. As has been noted in the literature [4], there have also been anecdotal claims by listeners that an improvement in fidelity can be noticed for sampling rates in excess of the 44.1 kHz sampling rate of the digital compact disk (CD) even though the listeners cannot hear pure tones above the 22 kHz Nyquist frequency. Such subtle effects may be masked in average mass-produced commercial audio systems and audiometric apparatus used in psychoacoustic research, because of the limited resolution of the equipment—the bottleneck then arises from the limitations of the apparatus rather than the ear.

Several past efforts are documented in the literature that investigated the magnitude τ of the smallest temporal feature in a stimulus that was just discernable. The threshold τ values determined in these past works were not surprising, in the sense that they were always slower than the time scale expected simply from the ear's bandwidth limit: i.e., $\tau > 1/\omega_{max}$. Furthermore, the types of temporal features studied in those works—such as silent gaps or iterated ripples—have no direct bearing on sound

reproduction since such distortions cannot naturally arise in an audio chain.

The present work reports the detection of a temporal delay shorter than any that has been previously published, and one whose threshold τ underceeds $1/\omega_{max}$. Furthermore, the method and type of the temporal feature—a disparity between spatial path distances from two loudspeaker drivers—is a distortion that can actually be manifested in a real-life audio setup, since most speaker systems consist of multiple drivers and even those that don't will exhibit a temporal spread because of the finite dimensions of the driver. Such arrival-time discrepancies can play an even greater significance in multi-channel surround-sound systems. The present result provides a scientific basis for the anecdotal claims by audiophiles that fidelity requires time response in the microsecond range, and provides a solid quantitative standard for assessing the deteriorating effects of temporal delays and smearing in an audio chain.

1 BACKGROUND

The central goal in high-fidelity sound reproduction is to reproduce a sound with sufficient accuracy such that the errors in all domains are below their thresholds of detectability. Setting aside stereo and spatial-localization aspects, monaural sounds can be perceived as different because of (a) different frequency components, (b) different levels of components, (c) different relative timings and rise times of components, and (d) different phases of the individual components. The first two differences are often collectively referred to as "spectral"; although, strictly speaking, these should be refered to as differences in the "amplitude spectrum" or "intensity spectrum", since all alterations, including those related to phase and time, can be described through changes in the complex Fourier spectrum. Except in simple linear systems, the inter-relations between frequency, amplitude, time, and phase are not straighforward. In audio systems, for example, a crossover can introduce a frequency dependent phase difference without physically delaying the onset of one frequency band with respect to another; however, the effect of having unequal listener-to-subwoofer and listener-to-satellite speaker distances is best described by an overall delay between the two frequency bands.

Similarly, the hearing mechanism treats phase and time differences on separate footings and errors in the two do not have equivalent consequences. In a linear circuit, a delay Δt in a sinusoidal signal is related to its phase shift θ through $\theta = 2\pi f \Delta t$. In the hearing process, the sound signal is decomposed into separate frequency channels through an array of sensory inner hair cells (IHCs) tuned to different characteristic frequencies (CFs) and arranged tonotopically along the basilar membrane in the cochlea. This tonotopically separated information is carried by auditory nerve fibers (ANFs) to the cochlear nucleus (CN). The nerve impulses along the ANFs follow the phase of their corresponding acoustic signals for frequencies up to about 4 kHz. However, the hearing mechansim largely abandons cross-frequency phase information leading to the famous Ohm's (second) law [5, 6] whereby the ear is not acutely sensitive to phase shifts between well separated frequencies (despite large differences in waveform shape). This fact is helpful in the design of frequency-cross-over circuits where phase differences between low-pass and high-pass outputs are important mainly to the extent that they affect the amplitude response [3]. Above 4 kHz, the ANFs respond approximately with a plateau of activity for the duration of the tone with no synchronization between the firing pattern and the phase of the acoustic signal.

While cross-frequency phase coherence is less important, time coherence is a different matter. The auditory system is very sensitive to the synchronicity in the onsets of different frequencies—as is well known, instrumental timbre becomes ambiguous if the onsets and decays of the notes are removed [7]. In the cochlear nucleus, fast responding octopus cells act as synchronous AND gates (with ∼60 inputs each) to converge coincident ANF signals from different frequency channels [8, 9]. These cells respond sharply to well timed multichromatic activity at the onset of a sound and thus their output is a gauge of the stimulus slew rate [10]. If the initial temporal uncertainity in the ANF signal is represented by a Gaussian probability-density function $f(t) = f_0 e^{-(t/t_0)^2}$, then t_0∼125 μs can be taken as a rough estimate of the initial temporal spread since the ANFs lose phase locking with the acoustic stimulus around 4 kHz and respond only to the positive half cycle [11, 12]. The probability for N signals to arrive simultaneously in order to excite an octopus cell is proportional to the product of the probabilities. This gives an output probability function $f'(t) \propto (f_0)^N e^{-(t/[t_0/\sqrt{N}])^2}$ with a reduced temporal spread of $t_0/\sqrt{N}$. Besides the initial convergence factor of N∼60 at each octopus cell, the octopus-cell outputs undergo additional convergencies at higher neural levels (e.g., in spherical bushy cells in the lateral leminiscus). It is not clear to what extent these convergencies boost N and improve the time resolution; however, it is clear that the maximum possible convergence factor cannot exceed the total number of IHCs, which is ∼4000. Thus for transient stimuli, the auditory system's temporal acuity τ may be estimated to be in the 2–16 μs range, taking $t_0/\sqrt{N}$ with N=60–4000. Notice that the value of this τ has very little to do with the high-frequency audibility limit f_{max}. As described earlier, the auditory signals originate from hair cells arranged such that the ones clos-

est to the entrance of the cochlea (basal end), sense the highest frequencies. With age, the high-frequency hair cells progressively perish and there is a corresponding recession of f_{max}. Since τ depends only on N and not f_{max} directly, it is not as sensitively affected by age. For example, a 50% drop in f_{max} from 18 kHz to 9 kHz corresponds to a loss of 1 out of 10 octaves of hearing, i.e., a drop in N from $\sim$4000 to $\sim$3600. This worsens τ by only 5% (i.e., $\sqrt{4000/3600}$). This may explain why elderly listeners, who have difficulty distinguishing consonants, can nevertheless detect extremely minor imperfections (presumably in the time domain) in high-fidelity playback systems.

Another apparent disparity between frequency and time-domain responses arises in the detection of jitter. It has been shown that temporal jitters as low as $\delta T \sim 0.1$ μs can be audibly discerned for tones consisting of high-frequency pulse trains [13]. Naively this may seem to imply that the ear can hear frequencies in the MHz range. However, whereas an exactly periodic pulse train contains only harmonics of the fundamental frequency, jitter introduces additional frequency content into this spectrum, which provides a cue for discrimination [14].

The effect of limiting the bandwidth of an audio reproduction chain has another consequence. No matter what the upper cutoff frequency or degree of misalignment, there is always some attenuation of frequencies within the audible band. Whether this will audibly affect the timbre or not will depend on the auditory system's threshold for distinguishing an intensity change (referred to as a just-noticeable-difference or JND). This sensitivity is again not closely related to f_{max} at all.

Finally, even in the case of steady tones, frequencies above f_{max} can affect perception when presented in sufficient strength. While the external and middle ear tend to filter out ultrasonic frequencies, some of this energy nevertheless does reach the inner ear, especially through bone conduction [15–19]. This ultrasonic energy may then stimulate audible-frequency channels either directly or through the generation of audible byproducts by the non-linearities in the ear's mechanisms [20, 21]. Such ultrasound, even when it is inaudible when presented by itself, has been speculated to modify the perception of timbre when superimposed on audible harmonics [22, 23].

Thus frequencies above f_{max}, while not usually considered audible when presented as pure tones, may nevertheless influence timbre and detract from audio fidelity if removed. As a result, the minimum response time of an audio chain that can degrade transparency can be much less than $1/\omega_{max}$.

Several psychoacoustic experiments have been conducted to probe the smallest temporal feature that can be discerned. Some of these experiments further attempt to separate out the roles of neurophysiological mechanisms that are isospectral in amplitude (relying purely on timing and phase) versus spectral mechanisms (relying on differences in the amplitude spectrum). As far as audio applications are concerned—for example determining the minimum digital sampling rate for achieving transparency—what matters is the minimum temporal error that can be discerned, not what specific cues our ear uses to tell the difference. Many of the past experiments have investigated the audibility of gaps in sinusoids [24–26] and noise [27–30] (the latter being isospectral), resulting in threshold τ values in the 0.2–10 ms range. In another experiment [31], a gap Δt was introduced within a pulse and listeners tried to distinguish a single pulse of width 20 μs from a pair of 10 μs pulses whose onsets had a relative delay of $dt = \Delta t + 10$ μs; the threshold for detection was $dt \sim 20$ μs and it was shown that the discernment is of spectral origin (narrow pulses have very broad spectra, and separating a pulse pair produces conspicuous spectral changes over an extended frequency range). Isospectral variants of this experiment [32, 33], where a pair of unequal pulses was compared with its time reversed version, found a threshold pulse separation of about 200 μs. One recent experiment [34] used iterated ripple noise (the stimulus consists of the iterated addition of copies of a noise signal that have been successively delayed so that the final result contains a periodic ripple) and found a threshold of $\tau > 12.5$ μs. Note that the temporal distortions (silent gaps and iteratively produced ripple) considered in the above experiments are ones that would not naturally arise within an audio chain and hence are not directly relevant for sound reproduction. Furthermore the threshold τ values observed are large and exceed $1/\omega_{max}$.

The present experiment investigates exactly the kind of temporal distortion that is manifested in a typical audio chain: one caused by a spatial misalignment between two loudspeakers. The resulting very slight spreading in the waveform also relates closely to temporal smearing caused by an audio component's finite relaxation time and bandwidth. Furthermore, the obtained threshold $\tau \approx 6$ μs is not only shorter than found in previously published literature but it also, for the first time, underceeds $1/\omega_{max}$. Unlike all of the aforementioned experiments, delays in this experiment are introduced mechanically, thereby avoiding spurious non-linear byproducts and switching transients that can sometimes be generated when the mixing, delaying, and gating of the signals is carried out by electronic/electrical means. The present result also casts light on possible physiological mechanisms and their relationship to accepted values of the sound-level JND published in the literature.

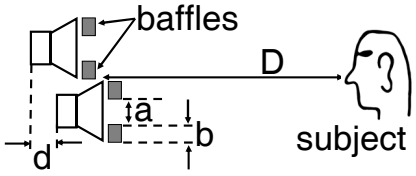

FIG. 1: *Experimental configuration. Speaker-to-listener distance D=4.3 m, aperture length a=1.5 cm, and speaker-center to speaker-center distance a + 2b = 9.9 cm. Misalignment offset d is variable. During blind trials, a listener tries to distinguish between the aligned (d=0) and misaligned (d≠0) settings for d values ranging 2–10 mm (τ∼6–30 μs).*

2 METHODS

2.1 Apparatus

The configuration of the experiment is shown in Fig. 1. Two loudspeakers are stacked vertically on top of each other with their front faces parallel to each other. The top speaker is mounted on rails and can slide back and forth between a fixed stop (for the aligned position) and a micrometer-setscrew adjustable stop (for the displaced position) through a set displacement d. The listener is seated at a distance $D = 4.3$ m, facing the speakers with ears at a height midway between the two speakers. The speakers are laterally centered w.r.t. (with respect to) both ears so that both ears receive the same signal.

The room shape is a rectangular parallelepiped with a height of 2.7 m, a width of 3.6 m, and a length of 5.8 m. The speaker-listener axis lies along the long dimension and is centered w.r.t. the side walls; this axis is at a height of 1.1 m above the floor. The floor and walls of the room were covered with acoustical carpeting and the ceiling covered with acoustical tiles. These materials have absorption coefficients of $S > 0.7$ at the frequencies of interest ($\geq$ 7 kHz). In addition, panels made from 38mm thick glass-fiber boards (for which $S > 0.95$) were placed at certain strategic locations to suppress principal reflections (there were six such panels with a total area of about 9 m^2).

The loudspeakers used were a pair of Aurum Cantus G2Si ribbon tweeters (Jinlang Audio Co. Ltd., Penglai City, P. R. of China) which have a frequency response of 2–40 kHz, a sensitivity of 96 dB/W at 1 m, and a nominal impedance of 6 Ω. Both speakers were connected in parallel to the same 7 kHz square-wave signal source. This signal source consisted of an analog signal generator (model 4001 manufactured by Global Specialties Instruments, Cheshire, Connecticutt) followed by a wideband amplifier (with a 3-dB power bandwidth of 0–2.2 MHz). Fig. 2 shows the voltage waveform at one speaker's input terminals measured with a LeCroy model LT322 (LeCroy Corporation, Chestnut Ridge, New York) 500 MHz digital storage oscilloscope, which digitized the signal at a sampling rate of 200 MS/s (million samples per second) and a 12-bit vertical resolution; this same oscilloscope was used in all the other waveform and spectrum measurements. Note the well controlled response with negligible ringing and overshoot, and rise/fall times of <0.2 μs. The measured jitter in this square-wave signal was 68 ns ($\lesssim$ 0.05% of the period).

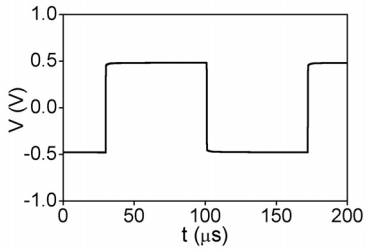

FIG. 2: *Waveform of voltage at loudspeaker terminals, recorded at a 200 MS/s sampling rate.*

A 7 kHz square waveform was chosen because it has only odd harmonics that, other than the fundamental, are below their respective single-tone audibility thresholds. However, the acoustic output from a loudspeaker will not be a perfect replica of its electrical input. Besides altering the harmonic coefficients because of an inconstant frequency response, subharmonics and other spurious anharmonic components may be generated [35] when a speaker is driven at high levels, especially with inadequate damping. In the present experiment the driving level is modest (∼0.5 V peak input voltage and 69 dB SPL sound level at listener position) and the damping is effective (<40 mΩ signal-source output resistance) to prevent anharmonic distortion. This absence of anharmonic distortion was verified by spectrum analyzing the acoustic output of the loudspeaker using an ACO Pacific (ACO Pacific, Inc., Belmont, California) model 7016 measurement microphone and a 4012 preamplifier with a 40 dB gain stage. The frequency response of the microphone together with its preamplifier was flat (±3 dB) within a 4 Hz–120 kHz band.

This power spectrum of the acoustic signal from the loudspeaker is shown in Fig. 3(a). Panel (b) shows a magnified view of the fundamental (f_1) peak with linear axes; the full-width half maximum (FWHM) is 0.77 Hz (∼0.01% of f_1). No subharmonic peaks could be distinguished from noise. The absolute sound level of the noise (including in the 3.5 kHz subharmonic vicinity) was $\lesssim$0 dB SPL (i.e., below the dashed line in Fig. 3[a]). There was also no detectable anharmonic content above f_1. As expected, the spectrum is dominated by odd harmonics (7, 21, 35, 49 kHz, ...) which extend well into the ultrasonic range. Because the entire signal chain is analog, spurious frequencies that can result from aliasing in dig-

ital systems were avoided.

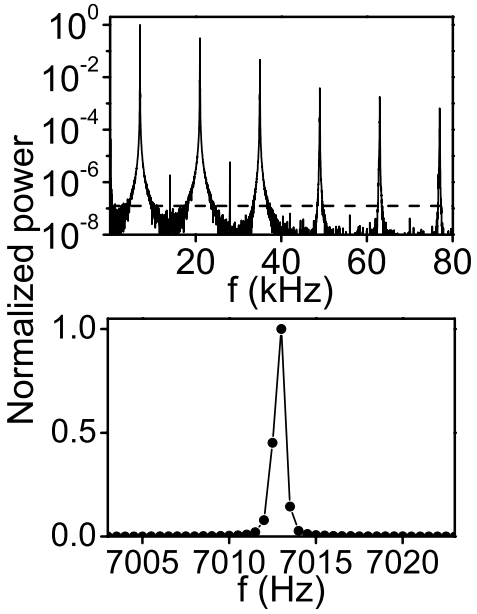

FIG. 3: *Power spectrum of the (unaveraged) acoustic output of one loudspeaker at a distance of 0.7 m. The power coefficients are normalized w.r.t. the fundamental peak. (a) Log-linear plot of the 20 Hz–80 kHz window in 20 Hz steps taken at a 2MS/s sampling rate. The horizontal dashed line corresponds to the absolute sound level of 0 dB SPL. (b) Linear-linear plot of the region near the fundamental peak in 0.5 Hz steps taken at a 50 kS/s sampling rate.*

2.2 Acoustic stimuli at the listener position

Referring back to Fig. 1, the two signals arrive at the listener's ears with a primary relative delay of

$$\tau = d/c, \tag{1}$$

where c is the speed of sound. As mentioned earlier, the listener is seated at a distance $D = 4.3$ m, facing the speakers with ears at a height midway between the two speakers. The alignment of the apparatus and listener position should be such that the line joining the midpoint between the listener's ears and the midpoint between the two speakers should be perpendicular to the plane defined by the speaker front surfaces when they are undisplaced. This was first checked with a laser beam and then the listener further fine tuned his/her head centering by ear. From this point onward the listener held his/her head still in this fixed position. The sound level at the subject location was 69 dB SPL for d=0.

Eq. 1 would be exact and complete if the two speakers were point sources and there were no room reflections. However, sound produced by an extended source will suffer an intrinsic temporal spread δt (because wavelets emanating from different points of the radiating surface will arrive at the ears at different times) which tends to reduce the sensitivity of the experiment. Here sound is generated in each tweeter by a 6 cm long vertical aluminum ribbon. As shown in Fig. 1, absorbent baffles were used to cut the aperture (and effective ribbon length) to a=1.5 cm (with b=4.2 cm), thus reducing the temporal spread to

$$\delta t \simeq (a^2 + 2ab)/2cD \simeq 0.5\ \mu\text{s}, \tag{2}$$

which is small compared with the delays probed in the experiment. The vertical separation of $a + 2b = 9.9$ cm, between the centers of the upper and lower speakers, results in an angular separation between these sources of 1.3^o at the listener position.

As with the temporal spreads within the sources, room reflections can also diminish the temporal definition of the delay. While precautions were taken to minimize reflected energy, even the best anechoic chamber will not have perfect absorption. Therefore it is necessary to quantitatively assess the effect of reflected energy on the experiment. The signal at the listener position will consist of the direct radiation from both speakers plus the sum of all reflections. Every reflected path originating from one speaker can be associated with a matching reflected path originating from the other speaker and all reflections can thus be considered in such matched pairs. Labelling each such reflection pair by n (with n=0 corresponding to the pair of direct signals), a speaker misalignment d introduces an incremental path difference d_n between the members of a pair. It is obvious from the geometry that $d_n \leq d$ in all cases. In fact for the first reflections from the sidewalls, $d_n \simeq dD/\sqrt{D^2 + w^2}$ (where w is the room width transverse to the speaker-listener axis) and for the first reflections from the floor and ceiling, $d_n \simeq dD/\sqrt{D^2 + (2h + b + a/2)^2}$ (where h is the vertical distance from the speaker-listener axis to the floor or ceiling). These d_n's correspond to delays $\tau_n = d_n/c \leq d/c$. The direct signal and the reflection from the back wall (behind the listener) both have $d_n = d$ and correspond exactly to the primary delay of d/c. Higher-order reflections will have progressively diminishing delays.

Besides having a shorter incremental delay between the two speaker signals, a pair of reflected signals can also have an extra initial geometrical path difference l_n between pair members. For paths that undergo a single reflection from either the floor or the ceiling,

$$l_n \approx [(a + 2b)D/2h] \left[\sqrt{\frac{D^2 + 4h^2}{D^2}} - \sqrt{\frac{D^2}{D^2 + 4h^2}} \right] \tag{3}$$

and for paths that undergo a single reflection from any

of the walls,

$$l_n = 0. \tag{4}$$

Each pair of reflected paths has an average path length D_n to the listener that increases with the order of reflections, with D_0=D for the direct signals.

With these definitions, the total signal at the listener position, from both speakers and summed over all reflections, can be represented (for each harmonic) by

$$A\cos(2\pi f[t + l'/c]) + A\cos(2\pi ft) = \sum_n A_n \cos(2\pi f[t + (D_n + l_n)/c]) + \sum_n A_n \cos(2\pi f[t + D_n]), \tag{5}$$

when the speakers are aligned, and by

$$A\cos(2\pi f[t + \{l' + d'\}/c]) + A\cos(2\pi ft) = \sum_n A_n \cos(2\pi f[t + \{d_n + D_n + l_n\}/c]) + \sum_n A_n \cos(2\pi f[t + D_n]), \tag{6}$$

when the speakers are misaligned (removing any overall constant phase that is common to all terms). The first terms on both sides of both equations correspond to the signal originating from the top speaker and the second terms to the signal originating from the bottom speaker. l' is the effective initial path offset and d' ($< d$) is the effective displacement (incremental path difference) averaged over all reflections through the above summations.

The amplitudes A_n fall off rapidly (w.r.t. A_0 of the direct signals) for paths undergoing multiple reflections. Even the first-order reflections are greatly attenuated w.r.t. the direct sounds because of three factors: (1) absorption at the reflecting surface, (2) longer path distance D_n and consequent inverse-square falloff, and (3) the narrow polar response of a ribbon tweeter which beams energy preferentially in the forward direction. This results in intensity (A_n^2/A_0^2) ratios of $\lesssim$2 % for the floor, ceiling, and side-wall reflections, and ∼20 % for the back-wall reflection. While the back-wall reflection can contribute to standing waves, it alters neither the offset nor the effective displacement (i.e., $l_n = 0$ and $d_n = d_0 = d$ for this reflection).

The path offset l' reduces the starting undisplaced sound intensity by the factor $\cos^2(\pi f l'/c)$. Displacing the speaker attenuates this starting sound level by the amount

$$\Delta L_p = 10\log\left[\frac{\cos^2\{\pi f(d + l')/c\}}{\cos^2\{\pi f l'/c\}}\right]. \tag{7}$$

If the listener's head is properly centered, l_0=0 for the direct signals; if the head is mispositioned by δy (estimated to be <3 cm) then $l_0 = \delta y(a + 2b)/D$. The other main contributions to l' come from the floor and ceiling reflections. From evaluating Eqs. 3 and 5, it can be estimated that the overall $l' < 1.2$ mm. From Eq. 7 one can now obtain upperbounds on the attenuations. These are given in Table I for the fundamental frequency.

TABLE I: *Theoretical upperbounds on signal attenuations for different displacements at the 7 kHz fundamental frequency.*

d (mm)	2.0	2.3	2.9	3.9	6.2	10.3
$-\Delta L_p$ (dB)	0.16	0.19	0.27	0.44	0.98	2.6

The preceding discussion and calculations were based on the geometry of the experiment and elementary signal theory. On the other hand, one can simply measure the actual final acoustic signal at the listener position with a microphone and analyze and quantify the changes. Fig. 4 shows the acoustic waveforms measured at the listener position for four displacements. The $d = 0$ curve corresponds to Eq. 5 and the other curves correspond to Eq. 6. The waveforms (measured using the aforementioned setup of an ACO measurement microphone and preamplifier followed by the LeCroy oscilloscope) each represent the average of 16,000 traces taken at a 20 MS/s sampling rate with a 12 bit vertical resolution.

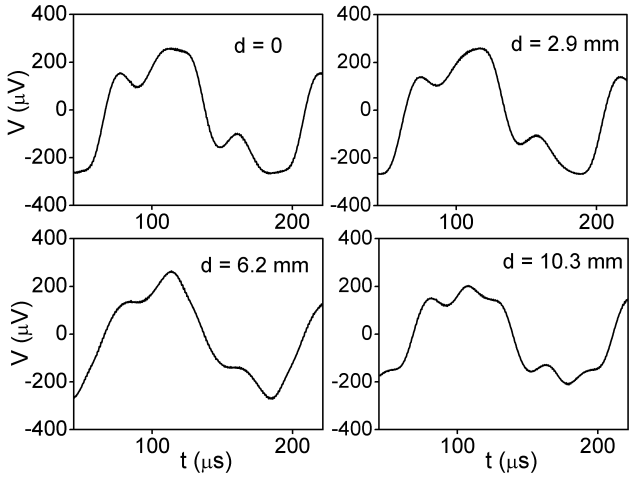

FIG. 4: *Acoustic waveforms measured at the listener position for four displacements. The $d = 0$ curve corresponds to the control stimulus (which is essentially twice the waveform produced by just one speaker). The waveforms were recorded at a 20 MS/s sampling rate.*

In view of the 7 kHz periodicity of the signals and absence of anharmonic components (as confirmed by the unaveraged spectrum of Fig. 3 and its earlier analysis) the waveforms of Fig. 4 can now be represented by a discrete Fourier series and are completely specified through the coefficients C_n and phases θ_n in the expan-

sion $V(t) = \Sigma C_n \cos(2\pi f_n t + \theta_n)$, where $f_n = n \times 7$ kHz. These coefficients are given in Table II and have been normalized w.r.t. the first-harmonic coefficient of the control waveform (i.e., by $C_1[0]$). Columns for harmonics where all C values fall well below the noise floor (0.005) have been excluded (this is the case for all even harmonics except for $n = 2$, which is on the borderline). The phase of each harmonic is specified relative to the phase of the fundamental for that same d value (i.e., w.r.t. $\theta_1[d]$); the absolute phase and the phase difference across different values of d are of course inconsequential. Besides the noise error, the values of the non-zero coefficients vary to some extent depending on the position of the microphone because of small local variations in intensity caused by weak partial standing waves. This positional sensitivity produces an uncertainity of about ±0.015 in the non-zero coefficients, which is equivalent to an error in the measured attenuations in the ±0.13 to ±0.16 dB range.

In a perfect square waveform, the harmonic coefficients are given by $C_n = 1/n$ for odd n and $C_n = 0$ for even n. It can be seen from Table II that the even C_n are indeed negligible and that the odd C_n follow $1/n$ well up to $n = 3$. Beyond that the high-frequency falloff of the speaker's response is evident and the coefficients are smaller than their $1/n$ theoretical values. The total sound level at $d = 0$ is 69 dB SPL. Most (88%) of this power is contained in the 7 kHz fundamental. The levels of all harmonics beyond 7 kHz at all d (for example at d=0, L_p[14kHz] ≈ 12 dB and L_p[21kHz] ≈ 60 dB) fall below their thresholds of audibility [35–37]. The last

TABLE II: *Harmonic contents of acoustic signals. Coefficients $C_n(d)$ are expressed as a fraction of $C_1(d=0)$. Phases θ_n, in radians, are expressed relative to the θ_1 for the same d value. The last two lower columns give the power attenuations, in dB, in the total rms strengths and first-harmonic components ($C_1(d)$) relative to their undisplaced ($d = 0$) control values. The noise floor and error bar for the coefficients are 0.005 and 0.015 respectively.*

d	f_1=7 kHz		f_2=14 kHz		f_3=21 kHz		f_5=35 kHz	
(mm)	C_1	θ_1	C_2	θ_2	C_3	θ_3	C_5	θ_5
0	1.000	0.00	0.006	-3.81	0.332	-1.29	0.145	-0.84
2.9	0.973	0.00	0.006	-4.09	0.329	-1.54	0.104	-1.66
6.2	0.933	0.00	0.005	-4.49	0.217	-1.50	0.038	-5.35
10.3	0.816	0.00	0.004	-5.48	0.156	-0.67	0.131	-0.11

d	f_7=49 kHz		f_9=63 kHz		f_{11}=77 kHz		Attenuation	
(mm)	C_7	θ_7	C_9	θ_9	C_{11}	θ_{11}	rms	C_1
0	0.018	-1.94	0.007	-3.02	0.002	-3.79	0	0
2.9	0.012	-3.09	0.003	-4.50	0.001	-0.10	0.26	0.24
6.2	0.013	-0.90	0.008	-2.62	0.002	-4.13	0.90	0.60
10.3	0.007	-1.98	0.005	-1.16	0.002	-2.98	2.03	1.76

two lower columns in Table II give the attenuations in the total rms power ($C_1^2(d) + C_2^2(d) + C_3^2(d) + ...$) and in the fundamental component (presumably the only audible Fourier component) relative to their values for $d = 0$. These measured attenuations in the fundamental component are plotted in Fig. 5 and are seen to more-or-less agree with the theoretical curve of Eq. 7 and corresponding calculated upperbounds of Table I.

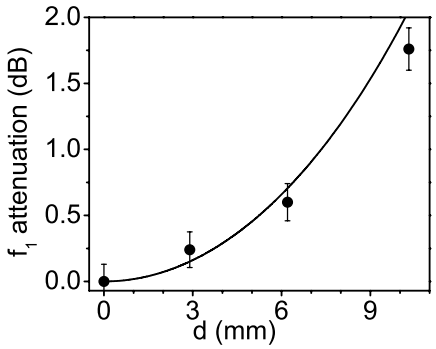

FIG. 5: *Power attenuation of the 7 kHz fundamental component as a function of speaker displacement. The symbols show the measured attenuation in the acoustic signal at the listener position. The solid line shows the theoretical curve corresponding to Eq. 7.*

Compared to the above measured acoustic waveform in air, the waveform at the eardrum will be poorer in higher-harmonic amplitudes because of filtering by the ear canal. So, for example, the third-harmonic to first-harmonic ratios at the eardrum will be lower than the measured $C_3(d)/C_1(d)$ ratios of Table II. However, the fractional change in each Fourier amplitude at the eardrum will be exactly the same as the corresponding measured $C_n(d)/C_n(0)$ ratio in Table II. Thus the attenuations in the fundamental at the eardrum will be exactly the same as the measured values given in the last lower column of the table and the total rms attenuation at the eardrum will be marginally lower than the second-last lower column of the table.

2.3 Procedure

In this experiment, subjects are seated in front of the two closely spaced speakers (Fig. 1). For the control condition, the speakers were aligned and equidistant from the listener's ears; for the test condition, the speakers were misaligned by a displacement d. The acoustic waveform at the ears becomes progressively temporally smeared (Fig. 4 and Eq. 6) and the harmonics increasingly attenuated (Eq. 7) as d is increased. The control sound ($d = 0$) was perceived to have a sharper or brighter timbre than the displaced setting ($d \neq 0$), until d became too small to make a difference. The goal was to find the threshold d that could barely be discriminated. In the blind test, the subject tries to judge whether an unknown sound corresponds to the control or displaced setting for different values of d. It was found that subjects typically need to listen to the sounds for a few seconds to form a

lasting impression of the timbre; quickly switching back and forth makes the discernment difficult.

The time course of the trials is as follows. For each d, a subject listens to the control and displaced sounds several times to become familiar with the timbre of each. The sounds are now played in the sequence: displaced, unknown, and control. The duration of the displaced sound 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 mechanical process of sliding the speaker from displaced to control position takes about 1 second and provides a smooth and continuous change from misaligned to aligned waveforms without the possibility of spurious transients or jumps in the waveform that can sometimes arise from an electrical switching method. The subject judges the identity of the unknown by comparing it to his or her recent memories of the known control and displaced sounds, after being allowed to listen to the entire sequence five times. Once the judgement has been recorded, the next trial for the same d setting is conducted. For each trial, the unknown sound is chosen to be either displaced 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 {0, 0, 1, 1, 0, 0, 0, 1, 0, 1}. When all ten trials for one subject have been completed, depending on the subjects' availability, either a new ten-trial set was conducted on the same subject for next lower setting of d or a set was conducted on another subject at the next lower d setting that he/she had not yet been tested for. A total of 50 blind trials was conducted over five subjects for each d setting. The tests were carried out for six values of d (2.0, 2.3, 2.9, 3.9, 6.2, and 10.3 mm) against the control value of d=0. Altogether this experiment consists of 300 blind trials.

2.4 Listeners

This study includes the participation of five listeners whose ages ranged 24–47 years. They had no history of hearing impairment or neurological disease. The subjects were volunteers and were not paid. The University of South Carolina Institutional Review Board (IRB) reviewed and approved the proposal for this research activity and the requisite participant consent forms.

3 RESULTS

Table III shows the results of the experiment. All subjects scored 100% on their blind tests for displacements down to d=2.9 mm. These scores of 50 out of 50 correspond to chi-squared values of $\chi^2 = 50$, well in excess of the critical value of 3.84. (A quantitative measure of discernability in psychophysical testing is the chi-squared analysis value defined as $\chi^2 = (C - T/2)^2/(T/2) + (I - T/2)^2/(T/2)$, where T is the total number of trials, C is the number of correct judgements, and I is the number of incorrect judgements. The critical value for χ^2 is 3.84 for a test with one degree of freedom—such as the present experiment where the speaker is judged to be aligned or not. Thus, for example, an 8 out of 10 score on a test would be considered statistically insignificant even though the percentage correct is 80%, since in this case $\chi^2 = 1.8 < 3.84$.) The shortest displacement that could be readily discerned was 2.3 mm, which corresponds to a delay of $\tau < 6.7$ μs. For this, combining all subjects, there were 82% correct judgements, a chi-squared analysis value of $\chi^2 = 20.48$, and a signal-detection-theory (SDT) discriminability index of $d' = 1.84$ with a criterion of $c = 0.97$. For d=2.0 mm, there was essentially no discernment between the displaced and control sounds (52% correct judgements, $\chi^2 = 0.08$, $d' = 0.14$, and $c = 0.18$).

TABLE III: *Results of blind trials. Each row corresponds to a different subject, arranged in order of ascending age. The entries correspond to the number of correct judgements (out of 10) for each subject for the indicated displacement d for that column.*

10.3 mm	**6.2 mm**	**3.9 mm**	**2.9 mm**	**2.3 mm**	**2.0 mm**
10	10	10	10	9	10
10	10	10	10	8	4
10	10	10	10	10	4
10	10	10	10	9	4
10	10	10	10	5	4

Eq. 1 relates the displacements to the primary delays between the two signals and Fig. 6 gives a graphical summary of the blind-trial results in terms of these delays. Panel (b) shows from chi-squared analysis that the demarcation between discernable and undiscernable delays lies around 6 μs.

4 CONCLUSIONS

The audio-reproduction chain contains many steps that can introduce errors in the time-domain, which can degrade sound fidelity: Every component's bandwidth limit (even if it behaves perfectly linearly) causes it to have a finite relaxation time of $\tau \sim 1/\omega_{max}$; use of digital carriers limits the shortest resolvable time interval to about half the sampling interval (which for CD would be 11 μs); and spatial dimensions of speaker drivers (or separations

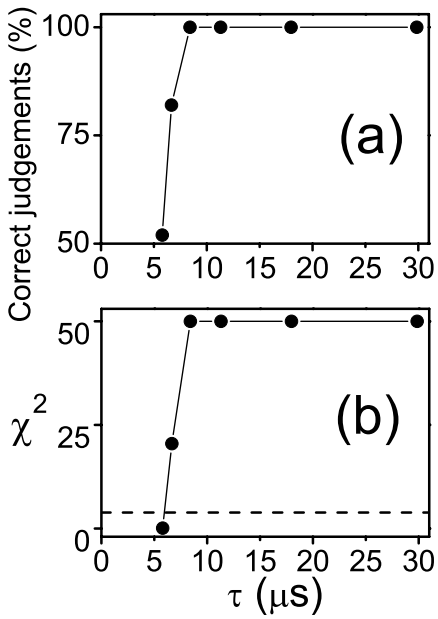

FIG. 6: *Summary of results as a function of the time delay (averaged across all subjects). Each data point consists of 50 blind trials. (a) The percentage of correct judgements. (b) Chi-squared value. The dashed line, corresponding to the critical value of 3.84, intersects the data curve around* $\tau \approx 6$ *μs.*

between multiple drivers) introduce temporal smearing and delays. In the last case, the temporal smearing can be enormous. For example, Eq. 2 indicates that a dipole loudspeaker with a single electrostatic panel of height a=1.5 m at a speaker-listener distance of D=5 m (with the listener's ear at half speaker height) will have a temporal spread of $a^2/2cD = 0.65$ ms. What this means is that even if the entire chain had an otherwise unlimited bandwidth, a delta-function (narrow impulse) input signal will get spread out over a 650 μs long rectangular window at the listener position. Thus a loudspeaker that subtends a large angle at the listener position must necessarily compromise fidelity, perhaps explaining why small speakers tend to have a subjectively cleaner and more coherent sound (although they may be deficient in their low-frequency response).

The present work provides the best current quantitative assessment ($\tau \approx 6$ μs) as to what extent such temporal errors make an audible difference. The vast majority of previous psychoacoustic experiments (summarized in the Background section) that probed this question, used equipment whose own own temporal response may have been a major limitation. Most of that research used rather coarse digital synthesis for the signal source, used amplification of insufficient intrinsic response speed, and transducers with limited bandwidths that were driven with inadequate damping. The present work uses an analog chain in which the square-wave signal presented to the transducer (including the response of both the signal generator and amplifier) has rise/fall times that are about 100 times faster than 48-kHz sampling-rate digital synthesis. The unsually low (40 mΩ) output impedance that sources the transducers provides exceptional damping and consequently a well controlled waveform, as shown in Fig. 2. The transducers used in this work have a far more extended bandwidth (spectrum shown in Fig. 3) compared with typical transducers used in audiometry (e.g., TDH-39 headphones). Thus by lifting some of the equipment bottlenecks, it was possible to demonstrate a much shorter threshold for discerning temporal errors, than has been achieved before (as noted in the Introduction section, the Δt defined in reference [31] corresponds to an interpulse delay of $dT = 10 + \Delta t \lesssim 20$ μs that is much longer than the threshold τ obtained here). This new lower threshold should be taken into account in the design and setup of audio components if the highest transparency is to be achieved.

While the present demonstration of discriminability at the microsecond time scale used simple (square-waveform) high-bandwidth signals, realistic musical sounds also carry content in this temporal-spectral range. Measurements of spectra of various musical instruments show that these extend into the ultrasonic range [38] and even beyond 100 kHz [39]. In the time domain, it has been demonstrated that several instruments (xylophone, trumpet, snare drum, and cymbals) have extremely steep onsets such that their full signal levels, exceeding 120 dB SPL, are attained in under 10 μs [2, 38]. Besides ultrasonic spectral content and microsecond-range onset durations, a third aspect of musical sound that demands fast temporal resolution is the reverberation. A transient sound produces a cascade of reflections whose frequency of incidence upon a listener grows with the square of time; the rate of arrival of these reflections $dN/dt \approx 4\pi c^3 t^2/V$ (where V is the room volume) approaches once every 5 μs after one second for a 2500 m^3 room [2]. Hence an accuracy of reproduction in the microsecond range is necessary to preserve the original acoustic environment's reverberation. The present experimental result thus provides a concrete basis for the anecdotal claims by audiophiles of sensitivity to very short time-domain errors (such as an insufficiency of some commonly used digital sampling rates) as discussed in the Introduction section.

While the neurophysiological basis that underlies the observed fast temporal discriminability is not of primary interest for sound reproduction, the present result nevertheless does shed some light on this issue. The starting point of any hearing sensation involves excitation of the inner hair cells in the cochlea. Three determinants can change the percept of a sound. One is a change in the stimulus frequency/ies, which changes the CFs and locations along the basilar membrane where IHCs are maximally excited. Another is a change in loudness, which changes the degree of excitation and the width of the

band of IHCs that are excited (the tuning curve of each IHC and associated ANF is about a third of an octave wide). These two "spectral" factors change the tonotopic excitation pattern. The third determinant affecting the percept of a sound is a change in the temporal order in which different IHCs and ANFs become excited. In this case the time averaged tonotopic excitation pattern will not change and presumably some kind of measurement or comparison of time takes place at neural stages beyond the cochlea. In the present experiment, the summed composite signal from the two speakers contains identical frequencies to each original signal, since no electronics or transduction is involved in the addition that could generate non-linear or anharmonic byproducts. How large are the sound-level differences and what is their possible role in the discernment? The frequencies present (Table II) in their order of (rapidly) declining intensities are 7 kHz, 14 kHz, 21 kHz, and 35 kHz (yet higher harmonics have intensities <1% of the fundamental). As per the earlier discussion in subsection 2.2, only the 7 kHz component exceeds its threshold of audibity. The sound-level changes in all components (individually or collectively) fall below their JNDs. For the shortest discriminable displacement of $d = 2.3$ mm, we have $\Delta L_p \approx -0.2$ dB (a 5% intensity decrease) for both rms and 7 kHz fundamental levels (Tables I and II, and Fig. 5). The JND (for $f \geq 7$ kHz and L_p=69 dB) is known from Jesteadt *et al.* [40] 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 the level changes in the experiment (< 0.2 dB) appear to be subliminal and the discrimination might involve more than just spectral-amplitude cues.

In the phase domain, it is worth comparing the present result with one by Plomp and Steeneken [41] where they investigate the distinguishability between low-frequency complex tones that differ only in phase but have identical amplitude spectra. In their experiment, one stimulus is composed of only sine (or only cosine) terms in the Fourier series and the other has alternating sines and cosines (i.e., a phase shift of 90$^{\rm o}$ between adjacent harmonics). For fundamental frequencies of 292.4 and 584.8 Hz, they concluded that the phase manipulated stimuli were not only distinguishable but that the audibility of the difference was equivalent to level changes of 2 and 0.7 dB/octave respectively. While their results cannot be straighforwardly extrapolated to 7 kHz, in the present experiment the phase shift of 15$^{\rm o}$ between the fundamental and the next prominent (third) harmonic, caused by the speaker displacement, is apparently equivalent to a level difference of ∼0.7 dB since it is just noticeable. It should be noted that the characteristic time differences between waveforms in reference [41] are large ($\gtrsim$100 μs) because of their lower frequencies.

5 POST-PUBLICATION NOTE

In a closely related experiment [42], signals were temporally smeared by passing them through a low-pass filter (to simulate the effect of bandwidth restriction in an audio component) instead of by spatially displacing speakers. In that experiment, the stimulus was conveyed to the subjects through supra-aural headphones instead of speakers. The threshold low-pass time constant that could barely be discriminated had a value of 5 μs, which is comparable to what was found in the present work.

6 ACKNOWLEDGEMENTS

The author gratefully acknowledges Mr. Gabriel Saracila for help with literature searching and other assistance, and acknowledges useful discussions and feedback on the manuscript from Professors James M. Knight, William M. Hartmann, Donata Oertel, Eric W. Healy, Fan-Gang Zeng, Raymond Meddis, Douglas H. Wedell, Alonso Botero, Kuniharu Kubodera, Varsha P. Kulkarni, and Antonello Monti. This work was partially supported by a grant from the University of South Carolina Office of Research and Health Sciences Research Funding Program.

* Electronic address: kunchur@sc.edu; URL: http://www.physics.sc.edu/kunchur

[1] H. R. E. van Maanen, "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 (1993).

[2] W. Woszczyk, "Physical and Perceptual Considerations for High-Resolution Audio", Audio Engineering Society Convention Paper 5931 Presented at the 115th Convention 2003 October 10-13 New York, New York (2003).

[3] N. Thiele, "Phase considerations in Loudspeaker Systems", Audio Engineering Society Convention Paper 5307 Presented at the 110th Convention 2001 May 12-15 Amsterdam, The Netherlands (2001).

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

[5] M. R. Schroeder, "Models of hearing", Proc. of the IEEE, **63**, 1332 (1975).

[6] W. M. Hartmann, "Signals, sound, and sensation (Modern Acoustics and Signal Processing)", AIP Press (1996).

[7] K. W. Berger, "Some factors in the recognition of timbre", J. Acoust. Soc. Am. **36**, 1988 (1963).

[8] D. Oertel, R. Bal, S. M. Gardner, P. H. Smith, and P.X. Joris, "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 (2000).

[9] N. L. Golding, D. Robertson, D. Oertel, "Recordings from slices indicate that octopus cells of the cochlear nu-

cleus detect coincident firing of auditory nerve fibers with temporal precision", J. Neurosci. **15**, 3138–3153 (1995).

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

[11] D. H. Johnson, "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 (1974).

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

[13] I. Pollack, "Submicrosecond auditory jitter discrimination thresholds", J. Acoust. Soc. Am. **45**, 1059–1059 (1969).

[14] I. Pollack, "Spectral basis of auditory jitter discrimination", J. Acoust. Soc. Am. **50**, 555 (1971).

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

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

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

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

[19] S. Fujioka et al., "Bone Conduction Hearing for Ultrasound", Trans. Tech. Com. Physio. Acoust. Soc. Japan, H-97-4 (1997).

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

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

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

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

[24] M. J Shailer and B. C. J. Moore, "Gap Detection and the Auditory Filter: Phase Effects Using Sinusoidal Stimuli", J. Acoust. Soc. Am. **81**, 1110–1117 (1987).

[25] C. Formby, M. Gerber, L. Sherlock, and L. Magder, "Evidence for an across-frequency, between-channel process in asymptotic monaural temporal gap detection", J. Acoust. Soc. Am. **103**, 3554–3560 (1998).

[26] B. C. J. Moore, "An Introduction to the Psychology of Hearing", 5th edition, Academic Press (2003).

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

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

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

[30] D. P. Allen, T. M. Virag, and J. R. Ison, "Humans detect gaps in broadband noise according to effective gap duration without additional cues from abrupt envelope changes", J. Acoust. Soc. Am. **112**, 2967–2974 (2002).

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

[32] D. Ronken, "Monaural detection of a phase difference between clicks, J. Acoust. Soc. Am. **47**, 1091–1099 (1970).

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

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

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

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

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

[38] P. Rogowski, A. Rakowski, and A. Jaroszewski, "Specific Hearing Loss in Young Percussion and Brass Wind Players due to Music Noise Exposures", The 8th International Congress on Sound and Vibration, Hong Kong, China, 2–6 July (2001).

[39] J. Boyk, "There's life above 20 kilohertz! A survey of musical instrument spectra to 102.4 kHz", http://www.cco.caltech.edu/ boyk/spectra/spectra.htm. Copyright ©1992, 1997 James Boyk, Music Lab, California Institute of Technology.

[40] W. Jesteadt, C. C. Wier, and D. M. Green, "Intensity discrimination as a function of frequency and sensation level", J. Acoust. Soc. Am. **61**, 169–177 (1977).

[41] R. Plomp and H. J. M. Steeneken, "The effect of phase on the timbre of complex tones", J. Acoust. Soc. Am. **46**, 409–421 (1969).

[42] M. N. Kunchur, "Temporal resolution of hearing probed by bandwidth restriction", Acta Acustica united with Acustica **94**, 594-603 (2008). (Preprint can be downloaded from http://www.physics.sc.edu/kunchur/temporal.pdf)