Hearing and Audio

Part 1 — Frequency, Phase, and Time

This two-part article series addresses controversial issues surrounding high-end audio and provides insights into the sophistication of the human hearing process. These concepts furnish a more enlightened backdrop for managing expectations, as to what might or might not matter for reproducing music to the highest fidelity.

By
Milind N. Kunchur

The realm of music reproduction known as high-end audio (HEA), takes a no-holds-barred approach for achieving sonic fidelity that is as close as possible to a live performance. The methods and materials used in HEA (e.g., atomic clocks, diamond diaphragms, etc.) can seem extreme and superfluous, and development of designs involves incremental changes that rely partly on sighted (i.e., not blind) listening tests. The minute and sometimes esoteric distortions that HEA seeks to minimize are beyond standard specifications. To what extent the claimed improvements can be heard is questioned because their audibility has often not been confirmed by controlled blind tests. For these collective reasons, HEA is surrounded by considerable controversy and skepticism. However, some of this skepticism is based on misconceptions about the link between time and frequency domains, and a lack of understanding of how hearing works. These concepts furnish a more enlightened backdrop for managing expectations, as to what might or might not matter for reproducing music to the highest fidelity. This discourse also explains why traditional quick A-B blind tests can fail. For further details, readers are referred to a recently published rigorous review article [1], as well as the other cited articles and articles posted on the author's homepage: (http://boson.physics.sc.edu/~kunchur).

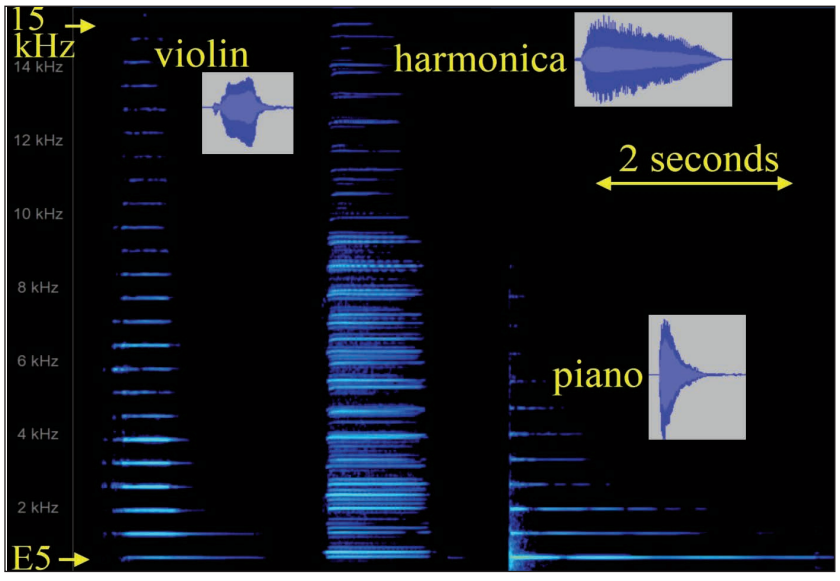


Figure 1: Spectrograms of the same musical note (E5) played on the different indicated instruments. The insets show the corresponding waveforms (voltage-versus-time graphs).

Relationship Between Frequency, Phase, and Time

Much of the audio community holds a misguided allegiance to the Fourier formalism and depends overly on the vocabulary of the frequency spectrum. For example, a component that doesn't alter the timbre [2] is often described as "neutral," implying a flat frequency response. "Accurate" or "natural" would be better adjectives because HEA components other than loudspeakers typically already have a flat enough frequency response. Distortions [3] affecting timbre are mostly related to the time-domain, such as the accuracy of attacks and decays of sounds, and not the frequency spectrum—sorry Fourier!

Figure 1 shows spectrograms of three instruments. The vertical axis represents frequency, the horizontal axis is time, and the brightness/color corresponds to the intensity. Attention is usually focused on the spectral energy distribution: the harmonica has the richest spectrum, jam packed with harmonics and other overtones; the piano has the purest tone with only the fundamental

surviving in the twilight of the decay; and the violin is in between. There are also clear time-domain differences in how frequency components individually grow and decay. For the violin, the fundamental begins a little later than some overtones. For the harmonica, the higher overtones have progressively delayed onsets. For the piano, all overtones start abruptly together. Furthermore, there is noise (fuzziness) between the piano's overtones at the onset, characterizing a sharp impulse [4].

Contrary to what some might believe, it is the time-domain differences not the spectral differences that most influence the timbre. This is illustrated in a demonstration video available on YouTube [5] (**Figure 2**). There is a dramatic change in timbre when notes are played backward in time even though the average spectrum, windowed over the entire waveform, is exactly the same [6]. This illustrates the importance of the envelope and the time domain for tonal quality.

The overwhelming importance of the delicate timing information associated with attacks/decays was strikingly demonstrated in the classic experiment by K. W. Berger [7] as illustrated in **Table 1**. Here various wind instruments were recorded and played back after marginally clipping off the beginnings and ends of the notes. Although the average spectrum was essentially unaltered, the removal of the onsets and offsets made the instruments unrecognizable even to the professional musicians who were well familiar with these sounds (e.g., six of them mistook the flute for an alto saxophone, five for a trumpet, etc.). These observations underscore the importance of time-domain performance in audio equipment. Indeed, many loudspeaker designs try to ensure that the acoustical centers of the drivers are equidistant from the listener, to correctly reproduce the timings between frequency components.

A time delay is sometimes confused with a phase shift. They are not the same and in general need not have any connection, except in special restricted

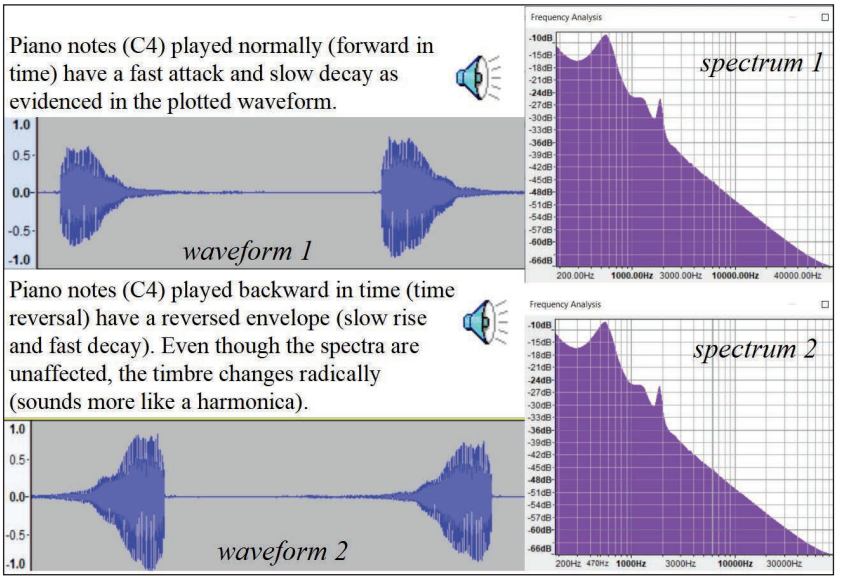
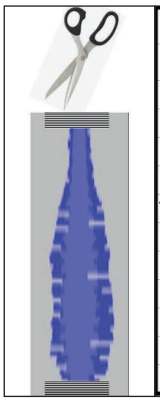


Figure 2: Demonstration [5] of timbre change with time reversal. The time-reversed piano sounds somewhat like a harmonica in timbre despite no alteration in the spectrum.

situations. If two people walk through a door, you can measure the delay between their arrivals with a stopwatch. It is meaningless to talk about a phase difference for such phenomena that are transient rather than oscillatory [8].

The spectrograms in Figure 1 show relative delays between onsets of the frequency components, not just phase shifts. This distinction between phase and delay is shown in a demonstration video [9] illustrated in **Figure 3**. The standard mix "A" (with matching phases and onset times) is played alternately with "B" (where the phases mismatch by 90 degrees): on a good audio system, little difference can be heard. Next "A" is alternated with "C" where the second harmonic's onset is time delayed (not just phase shifted); here a difference in the sound's character can be more readily heard.

Thus, a time delay between onsets is more serious than phase mismatches, and as mentioned earlier, many HEA loudspeaker designs pay careful attention to this. The timbral indifference to phase is such a widely established fact that there is a scientific law named for it: *Ohm's law of acoustics* [10, 11, 12].



Actual instrument ▼	Listener judgments									
	Flute	Oboe	Clarinet	Tenor sax	Alto sax	Trumpet	Cornet	French horn	Baritone	Trombone
Flute	1	2		1	6	5	4			4
Oboe		28								
Clarinet	1	1	20	4	3					
Tenor saxophone			25	2	1					
Alto saxophone				3	4		1	11	5	5
Trumpet	8				6	2	2	4	1	3
Cornet		1				12	15			
French horn	1			2	3			5	6	6
Baritone			1	1	2	3	2	4	7	3
Trombone	2	1		5	3			1	5	9

Table 1: Berger's "confusion matrix" experiment. Removing onset/offset transients and small changes in the envelope greatly obscured the timbre. Thus, spectral formants are not adequate for instrument identification.

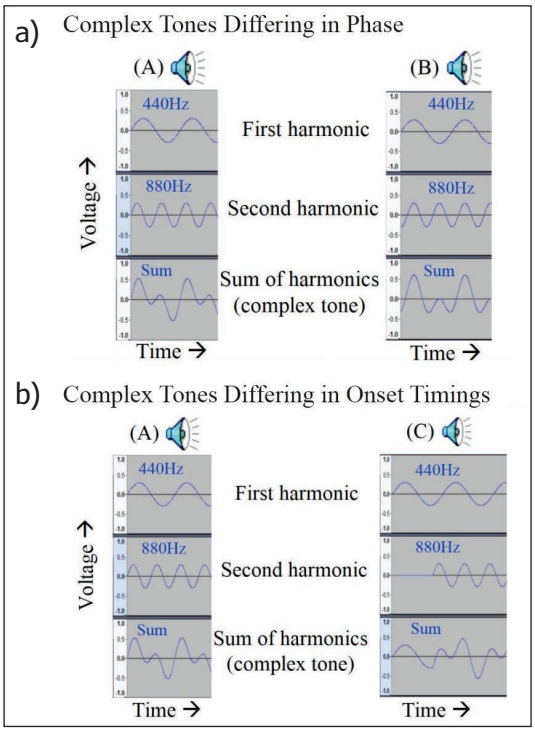


Figure 3: Demonstration video [9] shows that timbre changes for a time delay (b) between harmonics but not for a relative phase shift (a).

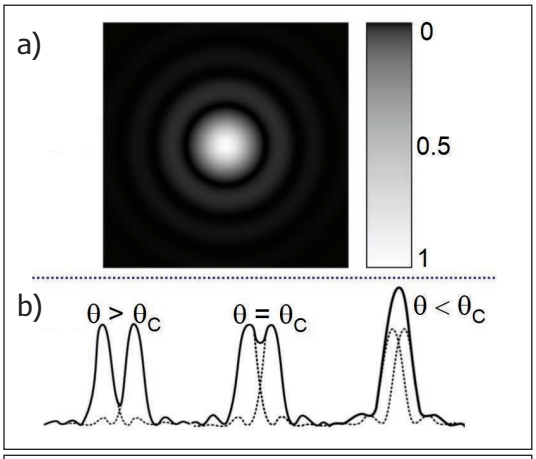


Figure 4: (a) A star images as a spread out "Airy" diffraction pattern. (b) Diffraction profiles of two close stars overlap and cannot be distinguished when their angular separation $\theta < \theta_C = 1.22\ \lambda/d$. (Figure adapted from cited Reference 1)

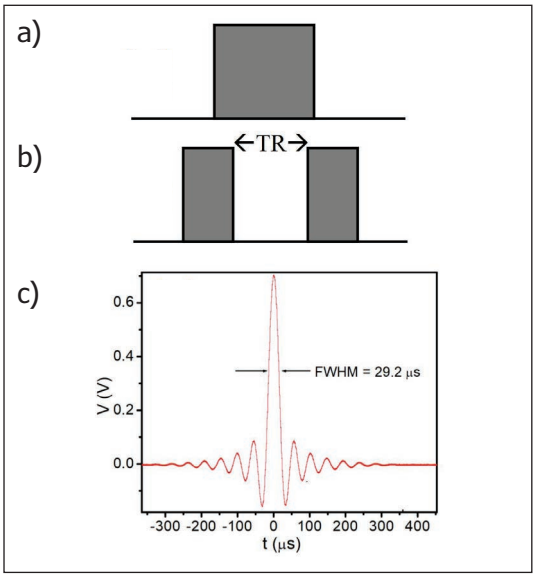


Figure 5: Experiments by Leshowitz [14] found that listeners could distinguish a single-pulse (a) from a double-pulse (b) separated by gaps of TR ~ 4μs–10μs. (c) the temporal smear (gauged by the full width, half maximum, or FWHM) of CD level (16 bits/44.1kHz) digital audio is considerably broader than this TR. (Figure adapted from cited Reference 1)

Temporal Smear and Resolution in Audio

Physical systems capable of a quicker temporal response are more likely able to produce and respond to higher frequencies. This is true, for example, when comparing a tweeter with a woofer: the higher resonance frequency goes hand in hand with a quicker return to equilibrium. Or an RC low-pass filter: a higher cut-off frequency $f_C = 1/2\pi RC$ implies a shorter characteristic decay time $\tau = RC = 1/2\pi f_C$. However, there is not always a straightforward reciprocal relationship between characteristic frequency and the characteristic time for a process. For example, reconstruction filters [13] in digital playback with the tightest temporal response can have a less extended frequency response. Daily events in our lives (e.g., eating breakfast) are "phase locked" to the 24-hour cycle. However, what does the duration of breakfast have to do with the 24-hour periodicity? Nothing!

Similarly, an individual does not need to have better high-frequency hearing to be able to resolve minute timing errors. The auditory binaural timing acuity is 10μs at 700Hz, which is a hundredth of the period T=1.43ms; also the acuity is 10 times worse (>100μs) at the higher frequency of 1400Hz! The interesting auditory neurophysiology underlying the disconnect between frequency, phase, and time, and the basis of Ohm's law, is explained in the second part of this article series.

The term "temporal resolution" gets used inconsistently to mean different things. For specificity, the term temporal smear τ will be used for the blurring of transient detail that is analogous to optical blurring in telescopes. **Figure 4a** shows the image of a star through a telescope: diffraction produces an "Airy pattern" instead of a single sharp point. Two close stars will merge into one when they are separated by less than the angle $\theta_C = 1.22\ \lambda/d$ radians (where d= aperture diameter and λ= wavelength), referred to as the Raleigh criterion. This is illustrated in the profiles of **Figure 4b**.

Similarly, every component in an audio chain temporally smears the signal and contributes toward broadening the audio system's overall τ. Two impulses closer than this τ will blur together. This can be expected to affect the sound quality if τ is greater than the transient resolution (TR) of our hearing. The classic experiments of B. Leshowitz [14] found TR ~ 4–10μs [15]. In analogy with the blurring of star images (Figure 4), this psychoacoustic study measured a listener's ability to distinguish single and double pulses of the same energy as illustrated in **Figure 5a**. This explains why digital audio at a 44.1kHz sampling rate ("Redbook" or CD level) is not fast enough. **Figure 5b** shows the voltage

signal coming out of a DAC [16] when the input signal (wave file) represents a single-sample impulse. τ is on the order of the sampling period and far wider than the auditory resolution. To tighten τ where it is no longer a bottleneck will require sampling frequencies considerably higher. However, most audio systems will have weaker links elsewhere in the chain. If the sampling rate is not the system's bottleneck, increasing it beyond 44.1kHz may not make a sonic difference.

Time-Shift Discrimination in Digital Audio

In digital audio there is a time $\tau^* \sim 1/[2^N f_S]$ that corresponds to the time-shift discrimination (i.e., the smallest time shift of a waveform that can be detected as a different digital value). **Figure 6** illustrates its meaning. For CD quality, τ^* is less than a nanosecond. [17] This very small value is sometimes misidentified as the "temporal resolution." It has no relevance for sonic fidelity. **Figure 5c** shows the mess that comes out for an intended infinitesimally narrow impulse—it is a hundred thousand times broader than the sub-nanosecond τ^* and the low-level contamination lasts even longer.

Resolution of Low-Level Detail

The temporal smear τ corresponds to the time interval over which the signal drops from its peak value to some fraction of order unity (e.g., ½, 10%, 1/e, etc.), such that two impulses within τ cannot be differentiated. However, if one looks at

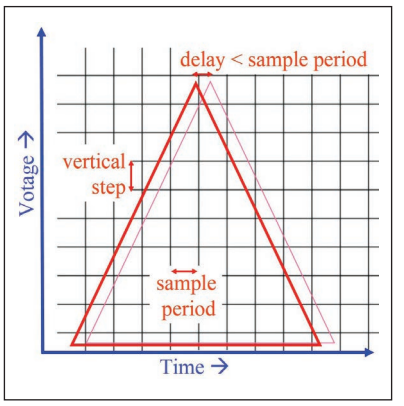


Figure 6: Time-shift discrimination in digital audio. A delay (time-shift) in a triangular waveform of less than the sampling period can be detected as a change in sample digital values. The detection threshold $\tau^* \sim [2^N f_S]\text{-}1$ is considerably less than the sampling period $1/f_S$, but is not indicative of the narrowest feature in the waveform that can be resolved. It is not "temporal resolution."

About the Author

Milind N. Kunchur is a Governor's Distinguished Professor and a Michael J. Mungo Distinguished Professor at the University of South Carolina in Columbia, SC. He is a Fellow of the American Physical Society and has won a Carnegie Foundation U.S. Professors of the Year award. He was named a Governor's South Carolina Professor of the Year and has received the George B. Pegram Medal, Ralph E. Powe Award, Donald S. Russell Award, Martin-Marietta Award, Michael A. Hill Award, Michael J. Mungo Award, and held a National Research Council Senior Fellowship. He has served as a panelist on High-Resolution-Audio and High-End-Audio workshops at Audio Engineering Society conventions.

Figure 5b, there is residue from the original peak that lingers much longer—on the order of a millisecond—that will contaminate subsequent sonic information. This will not be evident as a slowing of the attack, but it will obscure low-level detail. Its sonic effect can be expected to be worse than random noise because it is correlated with the signal. The depth of an image uses an auditory mechanism [1] that compares the direct intensity to the reverberant intensity (nominally 60dB lower after the reverberation time). Obscuring this information will reduce the depth and liveness of the reproduction. This extended smearing can be gauged by the decay cutoff time τ_C for the decay to disappear completely.

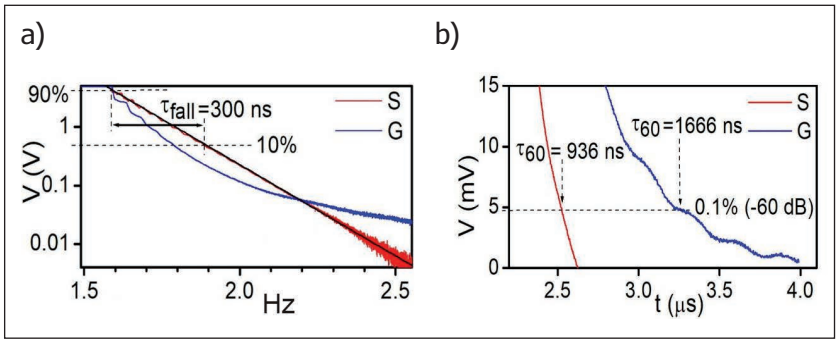


Figure 7: (a) Primary temporal smear (on the order of hundreds of nanoseconds) is more evident in the audiophile interconnect cable S than in the generic one G; however, the former has a clean cutoff in about a microsecond, whereas cable G has residue that lingers on for multiple microseconds. (b) τ_{60} is the time taken for the signal to drop by 60dB. (Data and figure adapted from cited References 1, 18).

Figure 7 shows the primary (τ) and extended (τ_C) temporal smearing in two interconnect cables. These parameters are somewhat independent of each other and may even reverse correlate. Another example is a well time-aligned loudspeaker with fast drivers but with a loose and poorly damped cabinet—this will have a short τ but a prolonged τ_C due to long lingering cabinet vibrations.

Currently standard specifications in audio do not report τ and τ_C. Perhaps these should be included in a future more comprehensive set of specifications. τ_C should be defined at the point when the residue from the impulse has become immeasurably small. As will be seen in the next part of this article series, the resolution of the ear is beyond astronomical—the number of variations supported by the cochlear neural excitation pattern has 17 more zeroes than all the stars in the universe!

References

[1] M. N. Kunchur, "The Human Auditory System and Audio," Applied Acoustics, Volume 211, pp. 109507 (2023), https://doi.org/10.1016/j.apacoust.2023.109507. Free preprint available at: https://arxiv.org/abs/2307.00084.

[2] Timbre (or tonal quality/color) is the quality that distinguishes notes with the same pitch, loudness, and duration.

[3] Except where specified, the term distortion will be used in the general sense to mean any alteration in waveform.

[4] A Dirac delta function contains all frequencies in its spectrum.

[5] M. Kunchur, "Timbre of forward and time-reversed piano notes," YouTube, https://www.youtube.com/watch?v=UIRuG4io2TM

[6] While a moving spectrum with a narrow window may provide some indication of what is going on, this is a rather indirect and convoluted way of representing what is more appropriately represented by a waveform and spectrogram.

[7] K. W. Berger, "Some factors in the recognition of timbre," *Journal of the Acoustical Society of America*, Volume 36, 1888 (1963).

[8] People tend to be proud about their knowledge of the Fourier formalism, but this can sometimes turn into a bad habit. Any signal can always be transformed from time domain to frequency domain. But sometimes instead of bringing out more meaning, it can obscure the essential information.

[9] M. Kunchur, "'Ohm's law of acoustics' and the effect of phase shift and time delay on timbre," YouTube, https://www.youtube.com/watch?v=qTDbHd_3EJU

[10] G. S. Ohm, "Uber die Definition des Tones, nebst daran geknupfter Theorie der Sirene und ahnlicher tonbildender Vorrichtungen," [English translation: On the Definition of Tone and an Associated Theory of the Siren and Similar Sound-Creating Devices] Ann. Phys. Chem., Volume 59, pp. 513-565 (1843).

[11] H. L. F. von Helmholtz, *On the Sensations of Tone*, English translation of 4th edition by A. J. Ellis (Longmans, Green and Co., London, 1912).

[12] For further discussion of Ohm's law, see section 3.8 of article [1].

[13] Sometimes also referred to as a smoothing, anti-aliasing, or anti-imaging filter.

[14] B. Leshowitz, "Measurement of the two-click threshold," *Journal of the Acoustical Society of America*, Volume 49, pp. 462–466 (1971).

[15] These experiments used equipment whose own temporal speed may have limited the precision. The experiments need to be repeated with modern HEA equipment to find the ultimate TR. The neurophysiological modeling described in the second part (Auditory resolution) of this article series suggests TR ~1μs may be possible.

[16] Muse Audio USB Mini DAC (other DACs and CD players tested differed in detail but had comparable FWHMs).

[17] With N = 16 being the vertical resolution (bit depth) and fs =44.1kHz the sampling frequency, we get: $\tau^* \sim 1/[2^{16} \times 44100] = 0.35$ns

[18] M. N. Kunchur, "Cable Pathways Between Audio Components Can Affect Perceived Sound Quality," *Journal of the Audio Engineering Society*, Volume 69, No. 6, pp. 398–409 (2021). DOI: https://doi.org/10.17743/jaes.2021.0012