

How human hearing is shaping high-end audio

Insights into how the auditory system processes time and information are guiding audio design beyond traditional measures of fidelity.

By **Milind N. Kunchur**

Designers of high-end audio systems strive to reproduce music with a realism that transports the listener to a concert hall by conveying the depth, height, and width of a musical performance. But from a physical perspective, certain design choices for those systems may seem extreme.

Some marketing materials, for instance, claim that listeners can perceive microsecond timing information in the sound with high accuracy and with little contamination from the equipment. The conventional upper frequency limit of human hearing, however, is 20 kHz. If you apply the familiar reciprocal relationship between time and frequency, that limit suggests that the human ear should be capable of perceiving time information with a maximum resolution of 50 μ s.

Such expectations of perception limitations arise in part from the extension of familiar physics concepts beyond the regimes in which they apply and from incomplete models of the neurophysiological mechanisms that underlie auditory perception. Indeed, the fine temporal and information resolution of human hearing is reshaping the understanding of how sound is perceived and how it needs to be reproduced to achieve the highest-fidelity listening experience.

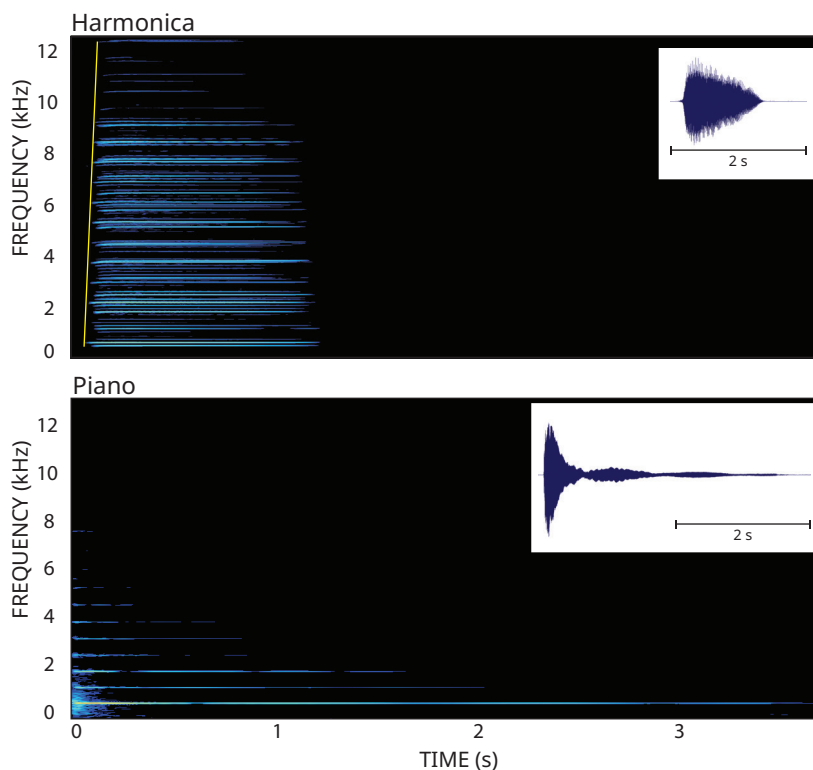


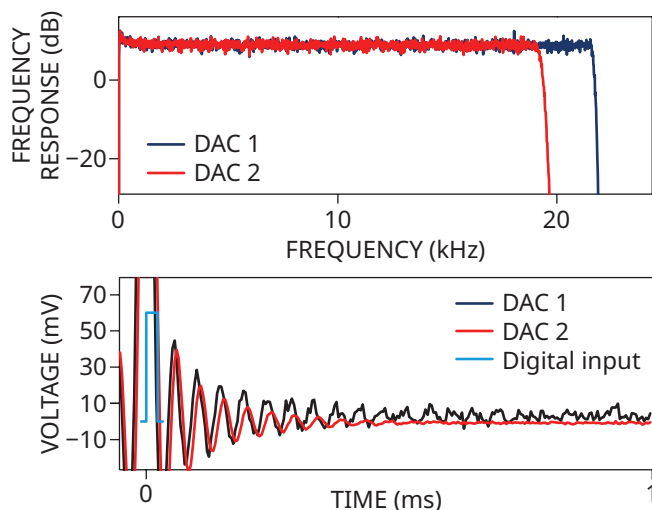
Figure 1. Spectrograms of a harmonica and a piano for the note E_5 show the time evolution of different frequency components of the sounds. The insets show the corresponding waveforms. The time-domain information in the plots is critical for how people perceive the different tonal qualities of the two sounds. The sloped line on the harmonica spectrogram is included to accentuate how the partial frequencies above the fundamental begin at progressively later times at higher frequencies.

Temporal precision

Audio engineers tend to think of musical sounds primarily in spectral terms. They evaluate equipment mainly using measurements such as the frequency response—the audio signals’ output-to-input ratio as a function of frequency—and distortions that modify the output spectrum. Those measure-

ments, however, do not provide a complete description of how people perceive sound.

Other properties of acoustic signals, particularly time-domain characteristics, are also important. A musical note is described by four perceptual attributes: pitch, duration, loudness, and timbre (its tonal quality). Reproducing pitch is relatively



▲ **Figure 2.** The frequency response (**top**) and impulse response (**bottom**) of two digital-to-analog converters (DACs). Although DAC 1 has a wider frequency bandwidth, its impulse response exhibits higher-intensity noise that persists for at least 1 ms. In listening tests, DAC 2 had higher perceived musical fidelity than DAC 1.

trivial—audio systems rarely struggle to get the notes right. The challenge is in faithfully reproducing timbre, and that’s where time-domain behaviors are especially pertinent.

Figure 1 shows spectrograms of a harmonica and a piano playing the note E_5 (the second E above middle C). Compared with the harmonica’s spectrogram, the piano’s has fewer partials, which are the harmonics and other frequencies above the fundamental of the note, here approximately 659 Hz. Certain temporal differences are more influential than spectral differences for the perception of timbre. The piano’s partials start almost simultaneously, whereas the harmonica’s partials begin at progressively later times at higher frequencies. As the waveforms in the insets show, the piano produces a more impulsive, faster-rising note than the harmonica does.

The importance of temporal structure for timbre can be illustrated by time reversing a note, which alters the timing without changing the time-averaged spectrum. The video “Time reversing a piano note” (see the online article for the link) demonstrates that a time-reversed piano note sounds like a markedly different instru-

ment—closer to a harmonica. The sensitivity to temporal structure is why high-end loudspeakers strive to provide microsecond synchrony. Multidriver loudspeakers can be time aligned, for example, so that the wavefronts from the different frequency bands of a sound arrive at the listener’s ear together.

So how does human hearing encode cross-frequency synchrony, which makes a piano sound like a piano, and with what precision? Let’s take a brief tour of the auditory system. Sound vibrations first enter through the outer ear and eventually reach the inner ear’s cochlea. There, the vibrations excite the basilar membrane, whose tapered structure gives rise to tonotopic tuning: Different frequencies are mapped to different positions along its length. About 3500 groups of hair cells distributed along the membrane convert the mechanical motion to electric signals that auditory nerve fibers transmit as spike trains to the brain. The arrangement acts as a 3500-channel spectrum analyzer.

Sharp, impulsive sounds, such as piano notes, produce bursts of activity across widely separated frequency channels. Those nearly simultaneous onsets are detected in

the brain by octopus neurons, which receive input from many auditory nerve fibers and function effectively as coincidence detectors. Octopus neurons are sensitive to the sharpness of temporal onsets, which is an important cue for perceiving timbres.

Neural analyses by myself and others suggest that a person’s transient auditory response has a resolution of about a microsecond. That’s consistent with several psychoacoustic experiments. The findings indicate that the pursuit of microsecond-level transient response—and the associated ultrasonic bandwidths that extend beyond 20 kHz—in high-end audio components may be justified, even if they are often regarded as unnecessary in conventional audio engineering.

Information resolution

In digital audio recording, standard 16-bit and 24-bit formats encode each sample at a rate of 44–192 kHz with 2^{16} (about 66 000) and 2^{24} (about 17 million) levels, respectively. Hearing, however, is not based on a single scalar measurement at each instant. Instead, sound is encoded in a distributed pattern of activity across roughly 3500 frequency channels in the cochlea and about 30 000 auditory nerve fibers.

Each cochlear channel senses a range of frequencies, and a channel’s range overlaps with about 100 neighboring channels. The overlap, which aids in capturing a sound’s dynamic range and temporal definition, can be represented approximately if you group the frequency spectrum into 40 equivalent rectangular bandwidths. Even if you adopt a conservative estimate of 10 distinguishable states per bandwidth, the number of possible instantaneous excitation patterns is about 10^{40} .

The vast information-resolving capacity of human hearing poses an extraordinary technical challenge for audio systems. In complex

sounds, such as a symphony performance, listeners can perceive subtle details, including the distinct timbres of dozens of individual instruments, their low-order reflections, and reverberation tails that linger from notes that were played seconds earlier.

The faint, temporally structured remnants of sounds are referred to collectively as low-level detail, and preserving it is critical for reproducing music with the realism of a live performance. In audio components, those subtle temporal features are often smeared by the equipment's lingering decays, such as post-transient ringing in loudspeakers and digital-to-analog converters (DACs), that persist after the signal has ceased.

To mitigate such decays, some high-end speakers have radiating components made of diamond or sapphire. Their high rigidity-to-mass ratios enable a faster and more controlled return to equilibrium. Some manufacturers now specify a cutoff time that characterizes when signals' ringing and other remnants decay completely. The transient response time and the cutoff time provide a more complete predictor of perceived fidelity.

Figure 2 plots the frequency and impulse responses of two DACs, which allow music stored in a digital format to be played through analog speakers. The first DAC has an extended frequency response and a slightly sharper impulse peak than the second, but it exhibits

residual ringing that persists for at least 1 ms. The second DAC shows a clean cutoff after 0.5 ms, which better preserves low-level detail. In tests, listeners perceived a higher-fidelity response from the second DAC, illustrating the importance of balancing frequency-domain performance with time-domain behavior.

Optimizing the transient response and the cutoff time of high-end audio equipment can lead to materials and designs that, at first glance, appear overengineered or superfluous. Yet those efforts help preserve the fine temporal structure of sound that underlies how the human auditory system perceives timbre and spatial information. The result is a convincing 3D auditory scene, rather than the illusion of instruments sitting on a line like birds on a wire.

PT

Additional resources

- ▶ M. N. Kunchur, "The human auditory system and audio," *Appl. Acoust.* **211**, 109507 (2023).
- ▶ D. Oertel, R. R. Fay, A. N. Popper, eds., *Integrative Functions in the Mammalian Auditory Pathway*, Springer (2002).

Milind N. Kunchur is a physics professor at the University of South Carolina in Columbia (www.kunchur.org).



aip.org

NOMINATIONS OPEN:

JOHN T. TATE MEDAL FOR INTERNATIONAL LEADERSHIP

**Celebrate Impactful Contributions to
Science and our Community**

Join the American Institute of Physics in
recognizing non-US scientists who have:

- Exhibited notable statesmanship on the international level
- Furthered international understanding and exchange
- Led in activities that elevate science in areas such as: publications, joint research projects, or successful international conferences



**Nominations close
Oct. 1, 2026**

**Scan to nominate a
colleague or yourself**