Audio Praxis

Hearing and Audio

Part 2 — Auditory Resolution

Our hearing, the conversion of sound into neural signals and the processing of these signals, is much more complex than many of us realize. The two-part article series addresses music reproduction and fidelity in that perspective. This second article addresses issues of resolution, detectability, and the consequences for blind testing.

By
Milind N. Kunchur

As mentioned in the first part of this article series, “Frequency, Phase, and Time,” some of the skepticism surrounding high-end audio (HEA) is based on a lack of understanding of how hearing works. This misapprehension leads to improperly designed blind listening tests and a self-fulfilling prophecy that can falsely confirm HEA claims to be invalid. The following discussion is a simplified and condensed extract from the rigorous review article [1]. For further details, serious readers are referred to this review article, as well as the other cited articles and articles posted on the author’s homepage (http://boson.physics.sc.edu/~kunchur).

Physiology of the Human Ear

Figure 1 shows a cross section of the human ear. After a series of nonlinear and actively controlled stages in the external and middle ear, the sound vibrations enter the liquid-filled cochlea—traveling up through the vestibular canal, making a U-turn at the far end (apex), and traveling back down the tympanic canal before reentering the middle ear. This traveling wave provokes vibrations in the basilar membrane, in which are embedded 3500 rows of hair cells. **Figure 2** shows a section of the basilar membrane. It is tapered such that high frequencies resonate maximally near its entrance (oval window) and low frequencies resonate maximally near its far end (apex). The structure acts as a nonlinear highly complex 3500-channel spectrum analyzer, and more. The voltage produced by the inner-hair cells (IHCs), which act like microphones, is influenced by five stages of active feedback mechanisms. Subsequent neural processing performs multiple spectro-temporal analyses, extracting correlations between the different channels such as the relative timings of onsets. For these reasons, a simple-minded spectral analysis of audio equipment hardly begins to predict its sonic performance.

Resolution of Detail

The 3500 “analog” (continuously variable) IHC voltages are represented by the “digitized” firing pattern of 30,000 auditory nerve fibers (referred to collectively as the neural excitation

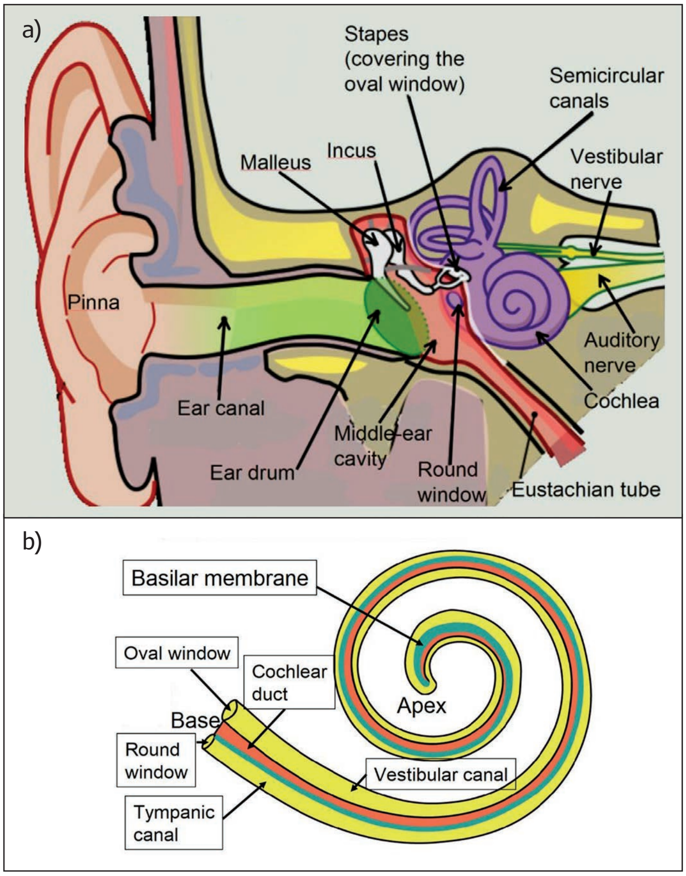

Figure 1: (a) Cross section of the human ear. In the external ear, sound enters through the pinna, traverses the ear canal, and is converted into solid vibrations at the ear drum. In the middle ear, these vibrations undergo mechanical and hydraulic amplification before transmission into the cochlea through the oval window. (b) Cross section of the cochlea. Vibrations enter the vestibular canal through the oval window, travel to the apex, and return back through the tympanic canal. This traveling wave vibrates the basilar membrane between the tympanic canal and cochlear duct (middle canal). Figure adapted from cited reference [1].

pattern or NEP). The instantaneous information this represents is so enormous that to describe it as "astronomical" is an insult to the ear. A very conservative lower bound [1] for the resolution of detail (number of variations) is RD $>10^{40}$—a number that has 17 more zeros than all the stars in the universe! [2] This information undergoes massive neural processing as indicated in **Figure 3**. This flow chart is shown here to inspire an appreciation for the incredible sophistication of human auditory neurophysiology; its detailed workings are explained elsewhere [1].

The vast detailed information of the NEP creates a problem for blind listening tests. The traditional quick AB tests are based on the rationale that having a long gap between A and B may cause the subject to forget the first sound. This is a mistake because the subject will then be relying on short-term memory, which lasts up to 15-30 seconds and can hold only about four items. This can work well when a single parameter is changed in a simple sound (e.g., differentiating the just noticeable difference (JND) in frequency for a pure tone). However, differences in timbre due to audio distortions involve myriad subtle changes in the neural excitation pattern. These are beyond the capacity of short-term memory. To absorb this detailed information requires listening to complex music for long periods with a "palate cleansing" break of a minute or longer between A and B. This resets short-term memory and invokes the durable and infinitely more detailed long-term memory. This extended-multiple-pass (EMP) listening approach led to the historic proof that cable pathways between audio components make an audible difference [3].

Resolution of Time

The first part of this article series discussed the misguided practice of taking the reciprocal of characteristic frequency, $1/f_C$, as the characteristic time τ. This doesn't always work even in physical systems and works even less reliably in hearing.

In simplified terms, there are two types of temporal resolutions operative in hearing—one is the (binaural) interaural time difference (ITD) between the two ears for a given frequency, the other is the (monaural) transient resolution TR, which is the timing difference between different frequencies for a given ear. These respectively follow the principal pathways AVCN→MSO and PVCN→VNLL shown in Figure 3.

We first consider the binaural ITD. As illustrated in **Figure 4a**, sound arriving from directly in front of a listener (an azimuthal angle of $\theta=0$) reaches both ears simultaneously; whereas for other directions it reaches with a relative delay of ITD = $d \sin(\theta)/v$. **Figure 4b** shows the classic results of published [5, 6] blind listening tests for the threshold detectable ITD at various frequencies. Note that the most sensitive detection of ITD (~9µs), which occurs at ~900Hz, is a hundredth of the period! The ITD is actually worse at higher frequencies (e.g., ~140µs at 1400Hz). This goes against the conventional expectation of associating shorter response times with higher frequencies.

Tonal (timbral) fidelity in audio involves mainly monaural transient resolution TR. We saw in Part 1 of this article (March 2024) that relative timings between onsets of different frequency components and attack (plus other stages of the envelope) are key. In the brain, there is specific circuitry that detects this cross-frequency synchronicity. The TR detection process begins with octopus cells (OCs) in the PVCN (see Figure 3). **Figure 5a** shows an image of an octopus cell. An OC's dendrites ("inputs") are

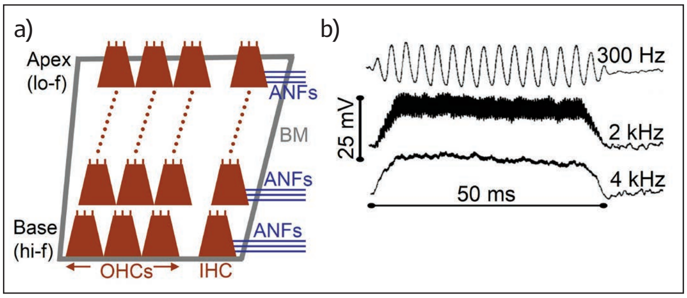

Figure 2: (a) A section of the basilar membrane showing some rows of hair cells. The inner hair cells (IHCs) work like "microphones" and produce a voltage in response to vibration. The outer hair cells (OHCs) work like "loudspeakers", producing sound and movement to actively enhance or dampen the response of the IHCs (kind of like active noise-cancellation headphones). Each IHC is tuned to a different frequency. The entire system acts like a 3500-channel spectrum analyzer. (b) The output voltage of inner hair cells [3]. The high-frequency channels produce a plateau for the duration of the sound—there is no cycle-to-cycle phase information, just timing information about the onset for those frequencies. Figure adapted from cited reference [1].

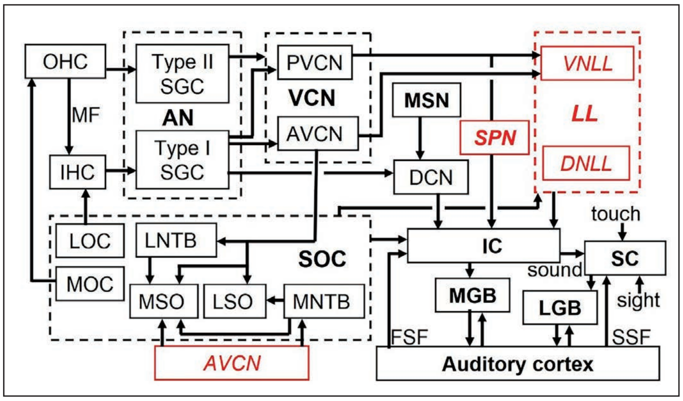

Figure 3: Analog (continuously variable) receptor voltages arising in the inner hair cells (IHC) become "digitized" and ascend upward through the auditory nerve (AN) to the posterior-ventral (PVCN), anterior-ventral (AVCN), and dorsal (DCN) cochlear nuclei. The pathway AVCN→MSO detects (binaural) interaural time differences (ITD) between ears per frequency. The pathway PVCN→VNLL detects (monaural) time differences (ITD) between frequency onsets per ear. The aforementioned processing takes place in the brainstem. Other pathways and overall neural analyses are explained in detail elsewhere [1]. Abbreviations: SOC = superior olivary complex; MSO = medial superior olive; LL =lateral lemniscus; VNLL = ventral nucleus of the LL; DNLL = dorsal nucleus of the LL; SGC = spiral ganglion cell; OHC = outer hair cell; MF = mechanical feedback; MSN = medullary somatosensory nucleus; SPN = superior paraolivary nucleus; LNTB = lateral nucleus of the trapezoid body; LSO = lateral superior olive; LOC = lateral olivocochlear system; MOC = medial olivocochlear system; IC = inferior colliculi; SC = superior colliculi; MGB = medial geniculate body; LGB = lateral geniculate body; FSF = frequency sharpening feedback; and SSF = spatial sharpening feedback. Figure adapted from cited reference [1].

perpendicular to the auditory nerve fibers, allowing it to integrate information from a broad range of frequency channels. Acting as a synchronous AND gate, the OC fires when its inputs arrive together within a tight time window.

Detailed neural analysis [1] estimates the resulting transient resolution to be TR ~ 1μs–10μs. This agrees well with the experimental threshold of 4μs–10μs determined in the psychoacoustic measurements of Leshowitz as discussed in Part 1. An important note is that TR is not closely tied to the highest frequency that an individual can hear. Thus, an elderly audiophile who cannot hear above a few kilohertz in the frequency domain may still be able to resolve temporal features in the microseconds range.

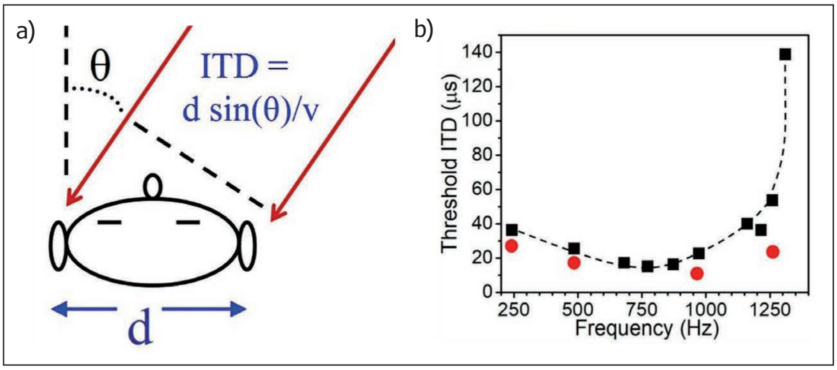

Figure 4: (a) Interaural time difference (ITD) for an off-axis sound is caused by a geometrical delay d sin(θ)/v, where θ = azimuthal angle, d=inter-ear spacing, and v=speed of sound. (b) Blind listening tests (psychoacoustic experiments) established [4, 5] that ITD is detected with microsecond precision. Figure adapted from cited reference [1].

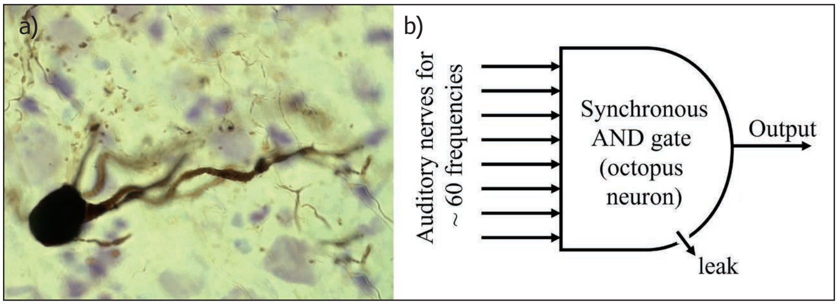

Figure 5: (a) An image of an octopus cell (a principal neuron of the PVCN). (b) This neuron acts as a synchronous AND gate—its leaky cell membrane ensures that the charge from each input decays quickly. The probability of an output spike reflects the overlap product of probabilities of input spikes from disparate frequencies, thus representing the sharpness of the sound transient. A detailed analysis [1] of the neural processing in the PVCN→VNLL pathway (see Figure 3), which includes parameters such as dendrite input resistances and cellular time constants, estimates a monaural transient resolution of TR ~ 1μs–10 μs.

Conclusion

This two-part article series provides a condensed overview of the sophistication, sensitivity, and resolution of the human auditory system. This provides a more informed background for discussing fidelity in reproduced sound, dispelling the many incorrect deeply rooted assumptions that permeate some portions of the audio community.

The serious reader is strongly encouraged to read the more detailed papers, which can be downloaded freely from the author's homepage. In particular, the aforementioned review on hearing [1] is an essential read for all who wish to properly understand what goes on inside our brains after sound hits our ears.

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. Professor Kunchur can be reached at: kunchur@gmail.com or kunchur@mailbox.sc.edu

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] The maximum value may be as high as RD ~ 10^{80}, which is roughly the total number of atoms in the universe!

[3] A. R. Palmer and I. J. Russell, "Phase-locking in the cochlear nerve of the guinea pig and its relation to the receptor potential of inner hair-cells," *Hearing Research*, Volume 24, No. 1, pp. 1-15. (1986). PMID: 3759671. DOI: 10.1016/0378-5955(86)90002-x.

[4] 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. Free preprint available under author's homepage http://boson.physics.sc.edu/~kunchur

[5] R. B. Klumpp, and H. R. Eady, "Some measurements of interaural time difference thresholds," *Journal of the Acoustical Society of America*, Volume 28, pp. 859–860 (1956). DOI: 10.1121/1.1908493.

[6] A. Brughera, L. Dunai, and W. M. Hartmann, "Human interaural time difference thresholds for sine tones: the high-frequency limit," *Journal of the Acoustical Society of America*, Volume 133, No. 5, pp. 2839-2855 (May 2013). Listed as Fig. 1(c). DOI: 10.1121/1.4795778. PMID: 23654390; PMCID: PMC3663869.

J SERIES

THE LEGENDARY J SERIES IS BACK!

TRIAD'S CLASSIC J SERIES TRANSFORMER DESIGN COMES OUT OF THE VAULT WITH ITS LEGENDARY PERFORMANCE PEDIGREE AND MORE. BORN IN THE 1950S, THE J SERIES TRANSFORMERS FEATURES HIGH-NICKEL LAMINATIONS, PAPER SECTION WINDINGS AND A SPECIALLY TOOLED MU-METAL SHIELDING ENCLOSURE. LONG RECOGNIZED BY ENGINEERS FOR ITS QUALITY SOUND, TODAY'S J SERIES IS IDEAL FOR NEW OR VINTAGE AUDIO GEAR, SOUND PROCESSING AND ANALOG STUDIO RECORDING EQUIPMENT.

- SHIELDED AUDIO IMPEDANCE MATCHING DESIGN FOR A WARM SIGNAL.
- VINTAGE CIRCUITRY, PREMIUM MATERIALS, HAND-CRAFTED QUALITY.
- SINGLE-HOLE MOUNTING ALLOWS ROTATION FOR MAXIMUM HUM REDUCTION.
- VISIT ONLINECOMPONENTS.COM TO ORDER OR LEARN MORE.

MADE IN THE U.S.A.

TRIAD
WWW.TRIADMAGNETICS.COM
PERRIS, CALIFORNIA, U.S.A.

onlinecomponents.com
THE ONLINE DISTRIBUTOR OF ELECTRONIC COMPONENTS