

THE SIGNIFICANCE OF THE COCHLEAR TRAVELING WAVE FOR THEORIES OF FREQUENCY ANALYSIS AND PITCH

STEVEN GREENBERG

*University of California, Berkeley and
International Computer Science Institute
1947 Center Street, Berkeley, CA 94704
steveng@icsi.berkeley.edu*

Auditory theory has traditionally pitted "place" (the tonotopically organized spatial pattern of excitation) versus "time" (the temporal pattern of discharge) with respect to the neural representation underlying specific attributes of acoustic sensation. This long-standing controversy has been of particular significance for models of pitch and frequency analysis, but casts its theoretical shadow over the discipline as a whole. A potential resolution of this historical opposition is proposed, in which place and time are viewed as flip sides of a complex representational matrix of neural activity, bound together through the mechanics of the cochlear traveling wave and its interaction with central loci of coincidence detection and inhibition. Frequency analysis is viewed as possessing two operational components. One is excitatory, based on spatially circumscribed patterns of temporally coherent peripheral activity and processed by central coincidence-sensitive neural elements. The other involves central inhibitory elements driven by non-synchronous activity distributed over a broad tonotopic domain. Together, these two mechanisms can account for the preservation of frequency selectivity across a wide range of frequencies and sound pressure levels, despite dramatic changes in the average-rate-based profile of neural activity. The traveling wave is also of importance in formatting the peripheral spatio-temporal response pattern germane to periodicity analysis and the perception of pitch. The present framework resolves the long-standing schism between spectral and temporal theories by virtue of a formulation in which pitch is viewed as resulting from the interplay of place and temporal information bound together into a coherent representation through the operation of central coincidence-sensitive neural populations. Within this perspective, both frequency resolution and neural synchrony are required for a robust sensation of pitch to occur.

1 Introduction

The displacement pattern of basilar membrane vibration is tonotopically organized, with high frequencies reaching their apogee towards the base of the cochlea and low frequencies achieving their maximum near the apex.¹ This systematic relationship between peak displacement and cochlear location serves as the linchpin of the "place" model of spectral representation and of auditory theory in general.

In recent years this "classic" place model has come under increasing scrutiny in light of experimental observations demonstrating that this spatial organization of excitatory activity is largely discernible only under a restricted set of conditions in the auditory periphery, thus calling into question its ability to subservise frequency coding at sound pressure levels typical of speech communication and musical performance.

In place of classic tonotopy many recent models of pitch and frequency analysis focus on the temporal properties of peripheral activity, principally the phase-locking behavior of single neural elements in the auditory nerve and ventral cochlear nucleus.^{2,3}

However, neither the temporal nor place approaches specify the operations through which the peripheral patterns are transformed into constellations of excitatory and inhibitory activity characteristic of the upper reaches of the auditory pathway, nor do they provide a principled account of the physiological basis for the perceptual stability of spectral representation and pitch extraction characteristic of human listening experience.

2 Cochlear Traveling Wave Delay

The motion of the basilar membrane proceeds in an orderly fashion from base to the point of maximum displacement, beyond which the wave damps out relatively quickly. The velocity of this traveling wave is very fast in the base, being nearly instantaneous (i.e., in the range of tens to hundreds of microseconds [μ s]) for frequencies above 4 kHz, but slowing dramatically for peak displacements in the apex. The transit time from base to apex requires 10 ms or greater.^{4,5}

The travel time for the cochlear traveling wave can be estimated from the latency of initial excitation of single auditory-nerve (AN) fibers to sinusoidal signals as illustrated in Figure 1. The cochlear delay (d_i) at a given frequency (f_i), can be modeled with a simple equation of the form:

$$d_i = f_i^{-1} + k$$

where k represents a delay constant of 0.002 s. This cochlear latency behavior represents the predicted response time of a minimum phase filter (f_i^{-1}) plus a 2-ms transmission time from transducer to the base of the cochlea. Although many properties of basilar membrane motion and filtering are non-linear, in terms of the cochlear traveling wave delay, the partition appears to act like a linear transducer.

There are two functional consequences of this latency behavior.

3 Latency Representation of Spectrum and Pitch

First, the shape of the delay function allows one to estimate the latency disparity between any two spectral components. For example, the initial spike latency for a 1-kHz signal is 3 ms, while that for a 0.5-kHz component is 4 ms, resulting in a latency disparity of 1 ms. Although this latency disparity is negligible for high-frequency signals, it can be considerable for low-frequency components within the core of the speech and musical range. It is also of significance that the initial latency

of excitation remains stable as a function of sound pressure level for auditory-nerve fibers whose characteristic frequencies are in close proximity to that of the signal. This important property of cochlear latency behavior is considered below (Section 4).

Such latency behavior provides a means of encoding low-frequency spectral information using a parameter of the peripheral excitation pattern distinct from, and yet still very much associated with the classical place mechanism. A given frequency produces a specific time signature that remains essentially stable over a circumscribed population of neurons. This latency signature is not only preserved, but appears to be enhanced at the level of the auditory cortex, where the latency disparities are magnified by a factor of three.⁶ There is some evidence that this latency behavior applies not only to the spectrum but also to encoding pitch-relevant information,⁷ as discussed below (Section 8).

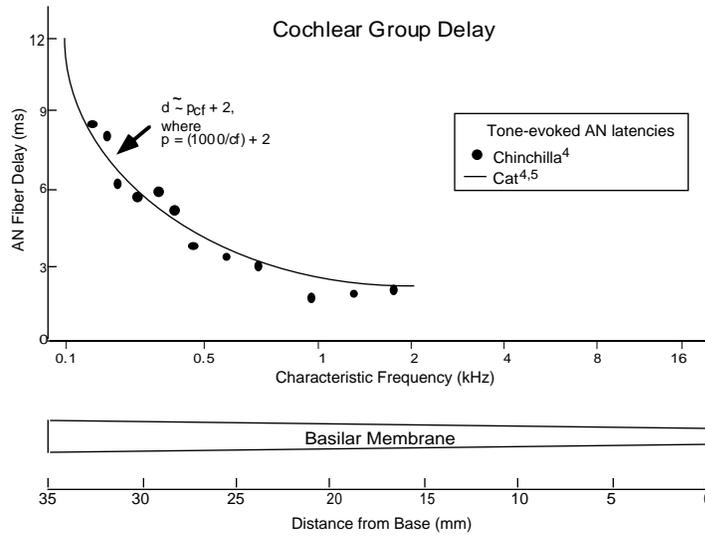


Figure 1 Latency of auditory-nerve fiber excitation as a function of sinusoidal signal frequency. Data derived from references 4 and 5. Solid curve fit to the equation indicated.

4 Stability of Spectral and Pitch Representation

The delay properties of the cochlear traveling wave also provide a potential means of stabilizing spectral information across a wide dynamic range of sound pressure levels and acoustic environmental conditions.

At the point of cochlear resonance, the initial latency of auditory-nerve-fiber excitation remains constant across sound pressure level.⁸ In consequence, a certain proportion of fibers will fire synchronously at onset in such a fashion as to potentially serve as a reliable frequency cue, particularly in concert with the latency gradient properties described above (Section 2).

Many neurons in the auditory brainstem and cortex appear to function as "coincidence" detectors, in that their probability of firing increases when multiple afferent input arrives within a narrow time window (typically on the order of tens or a few hundreds of microseconds). For example, at the level of the primary auditory cortex, cells typically discharge reliably only in response to stimulus onset and to certain forms of amplitude and frequency modulation, with firing rates rarely in excess of 20-30 spikes/s.⁹ The initial latency variance of such units is typically small (often less than 100 μ s).¹⁰ Although such cells are excitable by a relatively broad range of frequencies, the tuning of cortical neurons remains remarkably stable across sound pressure levels when measured in terms of discharge "doublets" and "triplets," which likely reflect large numbers of synchronous thalamic input.¹¹ The bandwidth of such multiple-onset tuning functions is roughly comparable to the critical band.

5 A Possible Latency Basis for the Critical Band

The critical band is a well-established limit of frequency selectivity that plays a central role in many perceptual attributes of sound.¹² Its physiological and anatomical bases remain obscure, although there appears to be a consistent relationship with distance along the cochlear partition (in the human, one critical band is roughly equal to 0.9 mm of cochlear length).¹³

A defining characteristic of the critical band is its relative constancy across sound pressure level.¹² In view of the appreciable broadening of tuning at high sound pressure levels observed in measures of selectivity based on average discharge rate, some other parameter of the neural excitation pattern is likely to form the basis of this frequency limit.

In view of the SPL-constancy of initial spike latency at cochlear resonance and the activation properties of coincidence-detector neurons, it appears likely that one important physiological component subserving frequency selectivity is the temporal coherence of peripheral neural discharge. Although the tonotopic extent of neural excitation increases with sound pressure level, the width of highly onset-

synchronized activity is unlikely to enlarge significantly. And since it is this highly synchronized activity that is particularly effective in exciting higher-level auditory coincidence detector neurons, it could readily serve as a reliable basis for encoding both frequency information and defining its limits of resolution. Thus, it may be that the physiological basis of the critical band may rest on the synchronicity of neural activity across a contiguous extent of a tonotopically organized gradient.

6 Possible Origins of Lateral Inhibition

Lateral inhibition serves to sharpen the spectral profile of excitatory neural activity.¹⁴ The origins of this important form of automatic gain control remains unclear. The magnitude and bandwidth of lateral suppression observed in the auditory nerve is considerably smaller than their counterparts driven by inhibition in many units (particularly choppers and pauser-buildups) of the cochlear nucleus.¹⁵

Both the bandwidth and magnitude of lateral inhibition among choppers and pauser-buildups increases with gain in signal intensity in a manner consistent with the tonotopic expansion of auditory-nerve activity as measured in terms of average rate.

The response properties of both choppers and pauser-buildups resemble integrators with relatively long time constants (for choppers, on the order of 2 to 10 ms, and for pauser-buildups, on the order of 20 to 100 ms).¹⁵ The precise timing of afferent input does not appear to be an important variable in determining their level of excitation.

The intensity-dependent nature of lateral inhibition suggests that its origins lie in the activity of neural elements relatively insensitive to input latency and which respond to the magnitude of activity largely irrespective of its temporal characteristic. Lateral inhibition thus appears to serve as a rough metric of integrated activity across a wide bandwidth of the tonotopic axis and to preserve a dynamic range of response for spectral coding.

7 Interplay between Synchronized Excitatory and Inhibitory Activity

The spectral representation of a sound may thus represent a combination of highly synchronous activity distributed over circumscribed tonotopic regions, signaling detailed information pertaining to aurally resolved components, along with a much more broadly distributed non-synchronized activity pattern associated with the overall energy level of the signal, and which serves as input to an automatic gain control mediated by neural inhibition.

8 Synchronous Activity as a Possible Basis for Pitch

Auditory-nerve fibers are capable of synchronizing their activity on a sustained basis to frequencies below 3 kHz.¹⁶ Circumscribed AN-fiber populations, synchronized to aurally resolved spectral components, provide a potential mechanism for encoding low pitch in the auditory periphery on the basis of "spectral" information associated with aurally resolved harmonics.

As a consequence of the temporal characteristics of the cochlear traveling wave described above (Section 2), the latency disparity among neighboring harmonics are such that the times of discharge associated with each spectral band are statistically correlated with those of its neighbors. As a consequence, it is possible to derive a periodicity analysis based on the correlation of synchronous activity across the spectrum in a manner resembling a running autocorrelation. The delay line, as it were, is derived from the cochlear traveling wave. The differential delays associated with specific periodicities appear to be magnified in primary auditory cortex, similar to the pattern observed for encoding spectral information⁸ (Section 3).

9 The Significance of the Cochlear Traveling Wave

There is much in our listening experience that cannot readily be accommodated within the current theoretical framework for auditory analysis of sound. Neither the classical place nor temporal models can, by themselves, account for many of the important properties of acoustic perception, such as the stability and apparent continuity of sound sources under a wide range of environmental conditions. A broader physiological and computational perspective is required to provide a more complete account of how such signals as speech and music are represented and processed in the auditory pathway.

Towards this end, it is suggested that the fine interplay between excitatory and inhibitory auditory processes may originate in the latency behavior of the cochlear traveling wave, providing the means through which different components of the acoustic signal are placed into appropriate temporal register for the central brainstem and cortex to operate on. In this sense the cochlear traveling wave may serve to temporally "preformat" the peripheral activity pattern entering the cochlear nuclei in such a fashion as to provide for fast, accurate and efficient extraction of features pertaining to pitch, intensity and spectral information. Thus, at least some portion of the transduction properties of the periphery (particularly those pertaining to spectral selectivity and filtering) are likely to be sculptured by requirements of the central auditory pathway.

References

1. Békésy, G. von (1960) *Experiments in Hearing* (McGraw Hill, New York).
2. Meddis, R. and Hewitt, M. (1991) Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I. Pitch identification, *J. Acoust. Soc. Am.* **98** 2866-2882.
3. Slaney, M. and Lyon, R. (1993) On the importance of time - a temporal representation of sound, in *Visual Representations of Speech Signals*, ed. M. Cooke, S. Beet, and M. Crawford (Wiley, Chichester) pp. 95-118.
4. Ruggero, M. and Rich, N. (1987) Timing of spike initiation in cochlear afferents: dependence on site of innervation, *J. Neurophys.* **58** 379-403.
5. Goldstein, J., Baer, T. and Kiang, N.Y-S. (1971) A theoretical treatment of latency, group delay and tuning characteristics for auditory nerve responses to clicks and tones, in *The Physiology of the Auditory System*, ed. M.B. Sachs (National Educational Consultants, Baltimore) pp. 133-141.
6. Roberts, T. and Poeppel, D. (1996) Latency of auditory evoked M100 as a function of tone frequency, *NeuroReport* **7** in press.
7. Ragot, R. and Lepaul-Ercole, R. (1996) Brain potentials as objective indexes of auditory pitch extraction from harmonics, *NeuroReport* **7** 905-909.
8. Kiang, N.Y-S. (1965) *Discharge Patterns of Single Fibers in the Cat's Auditory Nerve* (MIT Press, Cambridge).
9. Schreiner, C. E. and Urbas, J. V. (1986) Representation of amplitude modulation in the auditory cortex of the cat. I. The anterior auditory field (AAF), *Hearing Res.* **21** 227-241.
10. Clarey, J. C., Barone, P. and Imig, T. (1992) Physiology of thalamus and cortex, in *The Mammalian Auditory Pathway: Neurophysiology*, ed. A.N. Popper and R. R. Fay (Springer, New York) pp. 232-334.
11. Eggermont, J. J. and Smith, G. (1996) Burst-firing sharpens frequency-tuning in primary auditory cortex, *NeuroReport* **7** 753-757.
12. Scharf, B. (1970) Critical bands, in *Auditory Theory*, ed. J. Tobias (San Diego: Academic Press).
13. Greenwood, D. D. (1990) A cochlear frequency-position function for several species-29 years later. *J. Acoust. Soc. Am.* **87** 2592-2605.
14. Shamma S. A. (1985) Speech processing in the auditory system. II: Lateral inhibition and the central processing of speech evoked activity in the auditory nerve, *J. Acoust. Soc. Am.* **78** 1622-1632.
15. Rhode, W. S and Greenberg, S. (1992) Physiology of the cochlear nuclei, in *The Mammalian Auditory Pathway: Neurophysiology*, ed. A.N. Popper and R. R. Fay (Springer, New York) pp. 94-152.
16. Johnson, D. H. (1980) The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones, *J. Acoust. Soc. Am.* **68** 1115-1122.

