Presented at XIth Int. Symp. on Hearing, Grantham

A Space-Time Theory of Pitch and Timbre Based on Cortical Expansion of the Cochlear Traveling Wave Delay

S. Greenberg,[1] D. Poeppel[2] and T. Roberts[2]

University of California, Berkeley and International Computer Science Institute, 1947 Center Street, Berkeley, CA 94704, USA[1]

University of California, San Francisco, Department of Radiology, 513 Parnassus Avenue, S-362, San Francisco, CA 94143, USA[2]

1. Introduction

The displacement pattern of basilar membrane motion is tonotopically organized, with high frequencies reaching their apogee towards the base of the cochlea and low frequencies achieving their maximum near the apex (von Békésy, 1960). 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 generally discernible only under a restricted set of conditions in the auditory periphery, thus calling into question its ability to subserve 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 brainstem (e.g., Meddis and Hewitt, 1991; Slaney and Lyon, 1993). 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.

A new theoretical perspective, based on the latency characteristics of cochlear activity and its subsequent expansion at the level of the auditory cortex, is capable of accommodating both "place" and "time" within a unified representational framework that potentially accounts for the perceptual stability of auditory experience under a wide range of acoustic conditions.

2. The 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 it damps outs relatively quickly. The transit of this traveling wave is very fast in the base, its latency being nearly instantaneous (i.e., in the range of tens to hundreds of microseconds) for frequencies above 4 kHz, but slowing dramatically for peak displacements in the apex. The total travel time from base to apex may require 10 ms or longer (Ruggero and Rich, 1987; Goldstein et al., 1971), and 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_a) at a given frequency (f_i), can be modeled with a simple equation of the form:

$$d_a = f_i^{-1} + k \quad (1)$$

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

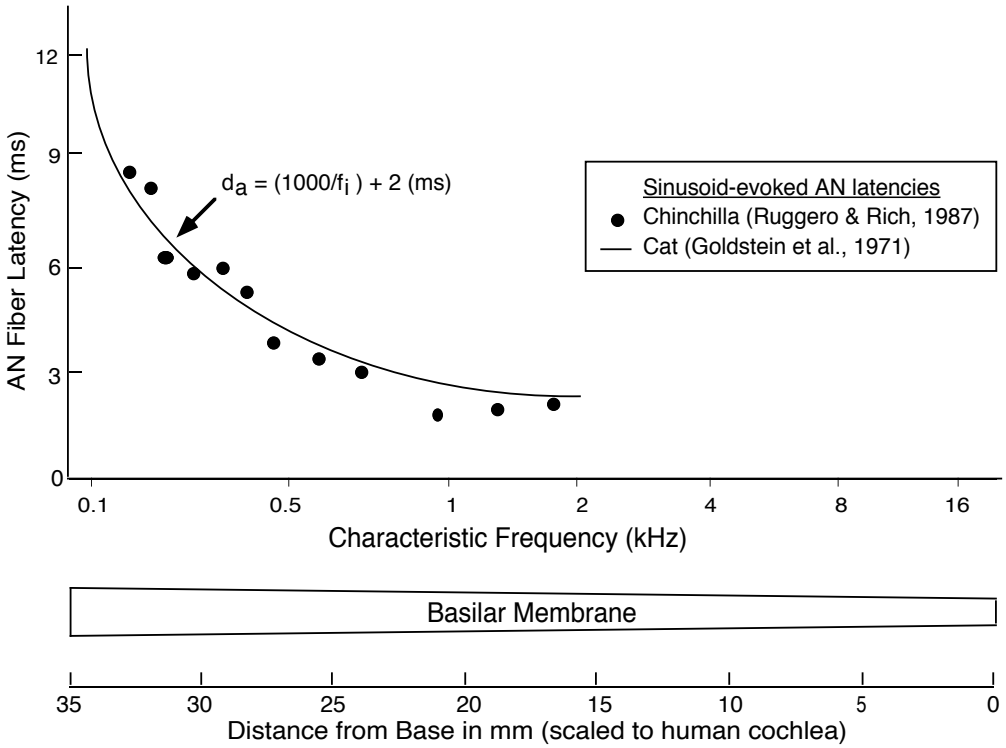

Figure 1. Latency of single auditory-nerve fiber excitation as a function of sinusoidal signal frequency. Data from Ruggero and Rich (1987) and Goldstein et al. (1971). The solid curve is derived from equation (1), where k=2 (ms). Adapted from Greenberg (1997). The human spatial-frequency coordinates are based on Greenwood's (1961) formulation.

3. Latency Representation of the Spectrum

The shape of the delay function allows one to estimate the anticipated latency disparity between any two spectral components. The latency disparity will be negligible (i.e., < 500 μs) for high-frequencies, but can be considerable for low-frequency components that lie within the core of the spectral range for speech and music. For example, the initial neural-spike latency in the cochlea for a 1-kHz signal is ca. 3 ms, while that for a 0.2-kHz component is ca. 7 ms, resulting in a latency disparity of 4 ms. Such latency behavior provides a potential means of encoding low-frequency spectral information using a parameter of the peripheral excitation pattern distinct from, yet still very much related to, the classical place principle of frequency coding. This frequency-latency association remains relatively stable across sound pressure level for those neural elements with characteristic frequencies in close proximity to the signal components (Kiang et al., 1965; Anderson et al., 1971), and thus provides a potential physiological basis for the representational constancy of the spectrum that characterizes much of our auditory experience (cf. Section 7 of this chapter).

This frequency-latency relation is preserved through the upper reaches of the human auditory brainstem pathway, as demonstrated in the latency behavior of the vertex-recorded, frequency-following response (Greenberg, 1980), an electrical-field-conducted potential reflecting the synchronous excitation of neurons in the inferior colliculus (Smith et al., 1975).

4. Cortical Expansion of the Cochlear Traveling Wave

At the level of the primary auditory cortex this frequency-dependent latency function frequently expands to between two and four times the initial cochlear delay. The magnitude of this expansion depends on a constellation of factors, including spectral shape and complexity, processing time (from signal onset) and (most likely) learning and attention. The specific nature of the latency expansion is not well understood. The present report describes some preliminary observations and relates these to a more general theoretical framework for processing complex sounds (such as speech and music) in the real world.

The basic latency function for the auditory cortex is shown in Figure 2, and conforms to equation (2):

$$d_c = nf_i^{-1} + m \qquad (2)$$

where n is the expansion factor (3 for the condition illustrated in Figure 2), and m is a transmission time constant of .1035 seconds. Details concerning the recording technique and stimulus generation/presentation are described in the appendix (and further discussed in Roberts and Poeppel, 1996).

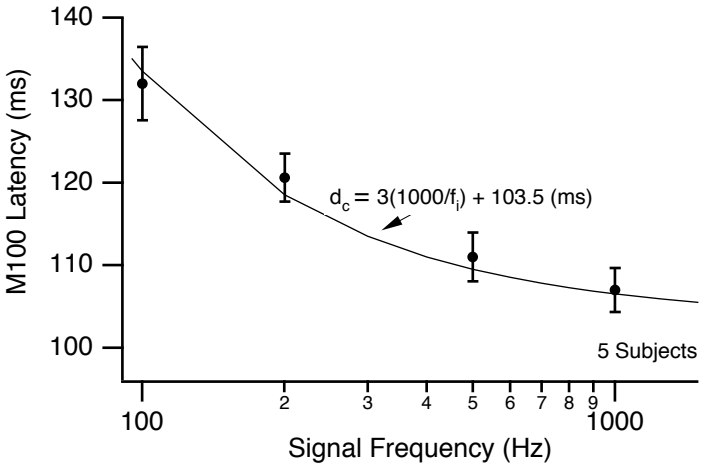

Figure 2. Latency of the M100 component of the cortical evoked-magnetic-field response averaged across five subjects. The standard error of the mean (error bars) for each experimental condition is also indicated. Stimuli were 400-ms sinusoids, presented at a sound pressure level of 60 dB, ca. once per second. Each experimental condition was based on the presentation of 100 repetitions of each stimulus. The frequency-latency function predicted by equation (2) is indicated by the solid curve(where n=3, m=103.5 ms).

5. Genesis of the Cortical Latency Expansion

The initial spike latency of most individual neurons in the primary auditory cortex is 17-25 ms (Dinse and Schreiner, 1996), just several milliseconds longer than the latency of single-unit activity recorded from the thalamic (medial geniculate body, e.g., Clarey et al., 1992) and pontine (inferior colliculus, e.g., Langner and Schreiner, 1988) levels of the auditory pathway. These neuronal latencies are mirrored in the response latency of specific components of the brainstem (Davis, 1976; Smith et al., 1975) and middle-latency, thalamo-cortical (Kraus and McGee, 1992) responses recorded from the scalp of human subjects.

Despite this early, first wave of cortical activation, primary auditory cortex fails to respond in a major fashion (as indexed by the magnitude of the evoked response) until ca. 100 ms following stimulus onset (ibid). Thus, there is ca. an 80-ms "gap" between initial activation (from the thalamic inputs) and the onset of mass, synchronized activity in the

auditory cortex (as reflected in the M100/N1 component of the evoked response). What happens during this "missing" 80 ms? And could the processes involved be associated with the expansion of the cochlear traveling wave delay?

5.1 Cortical Latency Expansion to Amplitude Modulated Signals

Relevant to a consideration of these issues is the latency behavior of the cortical evoked response to non-sinusoidal signals, as illustrated in Figure 3. Amplitude-modulated signals exhibit a smaller degree of latency expansion than evidenced by sinusoidal signals. The latency behavior of the cortical responses to AM and pulse-train signals can be modeled as conforming to expansion factors of 1 (i.e., no expansion) and 2, respectively. The significance of this variability in latency behavior is not well understood at present.

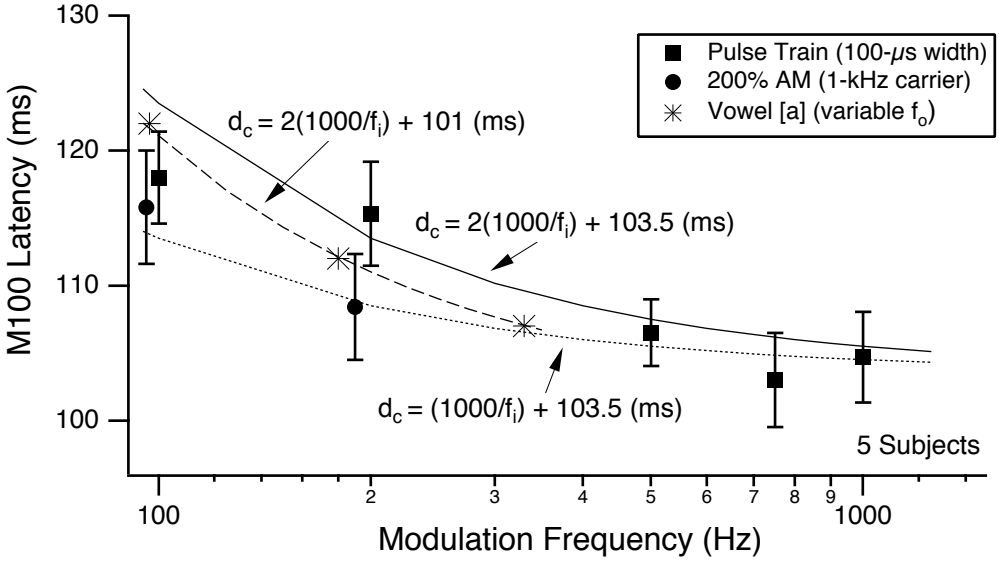

Figure 3. Latency of the M100 component of the cortical evoked-magnetic-field response averaged across five subjects in response to spectrally complex, amplitude-modulated signals. The solid squares pertain to pulse-train signals (100-μs pulse width), whose modulation rate varied between 100 and 1000 Hz. The solid circles pertain to 200% amplitude-modulated signals (with a carrier of 1 kHz, and equi-amplitude components at $f_C \pm f_m$). The solid curve conforms to equation (2), but where n (the expansion factor) = 2 and the transmission time constant, m = 103.5 ms. The dotted curve pertains to that instance of equation (2) where there is effectively no expansion (i.e., n = 1), conforming to the delay function derived from the cochlea (plus the 103.5-ms time constant). Sound pressure level = 40 dB SL. Data from Ragot and Lepaul-Ercole (1996) for single-formant stimuli are shown as stars and fit to equation (2) with n = 2 and m = 101 (ms). Data points for 100- and 200-Hz modulation-frequency conditions are slightly offset from each other for clarity of visualization.

Ragot and Lepaul-Ercole (1996) have observed a similar latency expansion (though they do not describe their data in this fashion) for single-formant, vocoid signals. In their study the latency of the N1 (the electrical analog of the M100) component of the cortical response systematically lengthens as the fundamental frequency decreases, as illustrated in Figure 3. The cortical latency behavior is consistent with an expansion factor of 2 and a transmission time constant of 101 ms.

Ragot and Lepaul-Ercole (1996) have also observed a concomitant shift in peak latency for the P2 component of the cortical response, as illustrated in Figure 4. The magnitude of the differential latency shift is approximately double that associated with the N1 component, resulting in an effective quadrupling of the cochlear traveling wave delay. These data imply the presence of a time-sensitive factor in the formation of the periodicity-latency cues at the cortical level.

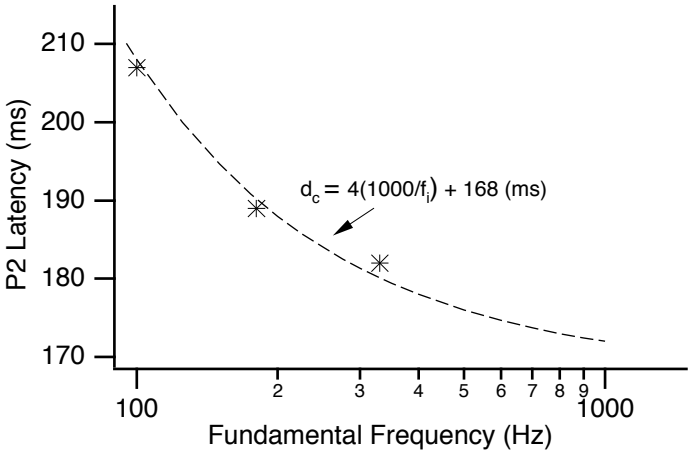

Figure 4. Latency of the P2 latency of the cortical evoked response as a function of fundamental frequency for the vocoid stimulus [a]. Data (based on an average of 8 subjects) from Ragot and Lepaul-Ercole (1996), and fit to a variant of equation (2), for n = 4, and m = 168 ms. Sound pressure level = 70 dB. Stimulus duration = 200 ms. Averages derived from ca. 200 repetitions of each stimulus.

6. Correlational Mechanisms Underlying the Latency Expansion

The latency pattern of the cortical evoked response behaves as if it were the product of a cascaded series of neuronal correlations. In the instance of sinusoidal signals, with a clearly defined pitch (and timbre), the latency function can be conceptualized as the result of two separate correlational operations, each of ca. 40 ms duration, that preserve and accumulate the frequency-dependent delay derived from more caudal stations of the auditory pathway. These neuronal operations are sensitive to the tonotopic extent over which the correlations are performed. The 40-ms quanta likely represent the interval over which local features are extracted from analysis of local channels. Spectrally more global operations require longer intervals to compute, representing the output of cascaded, short-term correlations taken over progressively broader tonotopic domains up to a limit of ca. 200 ms.

Figure 5 schematically illustrates the nature of the correlational operations envisioned. For spectrally local computations, such as those associated with a simple amplitude-modulated signal, the latency function will show little, if any, expansion beyond what is observed in the cochlea and brainstem. Broadband signals, such as pulse-trains and vowel-like sounds, will necessarily invoke cross-spectral correlational processes, as will low-frequency sinusoidal signals whose excitatory (phase-locking) shadow is cast across a broad tonotopic range at moderate to high sound pressure levels (Jenison et al., 1991).

7. Is Cortical Correlation the Basis for Perceptual Stability?

Neurons in the auditory cortex typically discharge at rates no greater than 25 spikes/s (Schreiner and Urbas, 1986), significantly lower than the discharge rate of afferent thalamic input, suggesting, in tandem with anatomical reconstruction studies (Winer, 1992), that the major input (and hence computational influence) to any given cortical cell stems from the auditory cortex itself. In this sense, the auditory cortex behaves like an inertial system, sampling over 40-ms quanta, in an effort to construct a stable, coherent "picture" of the external world that is in register across the sensory modalities and the motor system.

Such salient properties of our auditory experience, such as (1) the minimum duration for stable sensations of pitch and timbre (ca. 100-200 ms), (2) loudness integration (200 ms), (3) the prevalence of voice pitch in the range of 80-400 Hz and (4) the essential quantal (40-ms) nature of speech (Dudley, 1939) are commensurate with the current theoretical framework. Frequency- and periodicity-dependent latency cues provide the sort of stable neural representation required to account for the perceptual stability of our sensory world, and appear to be enhanced as a consequence of behavioral learning (Schreiner et al., 1997) and focal attention on a perceptually relevant task (cf. Langner et al., 1996).

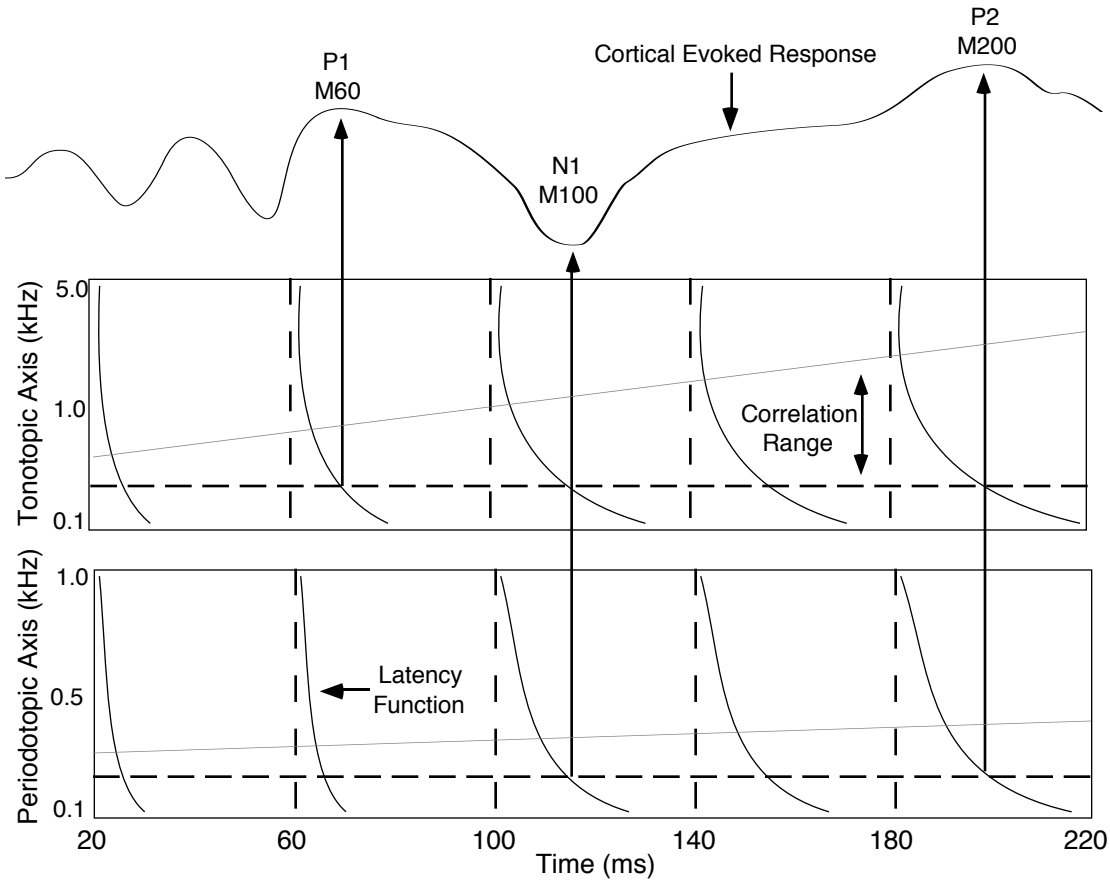

Figure 5. Schematic model of the mechanisms underlying the cortical expansion of the frequency/periodicity-latency function. Correlation of synchronous activity over specified tonotopic or periodotopic domain results in accumulation of latencies across 40-ms temporal quanta, resulting in expansion of the initial cochlear-derived function. Expansion operates at a longer time constant for the periodotopic domain, resulting in a smaller expansion factor than occurs for the tonotopic domain. Presumed cortical evoked response correlates of the expansion are indicated.□□□□□□□

8. Appendix - Stimulus Presentation and Recording Methods

Cortical latency data were derived from magnetic-field recordings using a 37-channel biomagnetometer (Magnes, Biomagnetic Technologies, Inc., San Diego) placed within a specially shielded, sound attenuated chamber. Subjects were healthy individuals with normal hearing thresholds. Stimuli were generated by a Wavetek (#395) function generator and presented to the right ear of the subject via an EarTone™ transducer specially designed for presenting acoustic signals in magnetically shielded environments. The sound pressure of the signals was set to either a constant level of 60 dB or to a level equivalent to 40 dB above the subject's threshold (SL) for each signal frequency. Signals were 400 ms in duration, presented at a rate varying, in pseudo-random fashion, between 0.7 and 1.3 times per second. Each stimulus was presented 100 times, and the resulting evoked-magnetic-field responses averaged and bandpass filtered between 1 and 20 Hz. The biomagnetometer array was positioned over the left temporal lobe, contralateral to the ear of stimulation. Single-equivalent-dipole modeling of the M100 component was performed in order to pinpoint the anatomical locus of the magnetic-field activity. These spatially defined patterns were set in register with high-resolution, 3-D magnetic resonance images for precise localization of the evoked activity. The single-equivalent-dipole model localized the magnetic field activity to the supra-temporal plane of the left hemisphere normally associated with primary auditory cortex. No clear-cut evidence of spatial-tonotopic organization was observed in this passive listening paradigm (cf. Langner et al., 1996).

9. References

Anderson, D. J., Rose, J. E. and Brugge, J. F. (1971) Temporal position of discharges in single auditory nerve fibers within the cycle of a sine-wave stimulus: Frequency and intensity effects. J. Acoust. Soc. Am. 49, 1131-1139.

Békésy, G. von (1960) Experiments in Hearing, McGraw Hill, New York.

Clarey, J. C., Barone, P. and Imig, T. (1992) Physiology of thalamus and cortex. In: The Mammalian Auditory Pathway: Neurophysiology, Popper, A.N. and Fay, R. R. (eds.), Springer, New York, pp. 232-334.

Davis, H. (1976) Principles of electric response audiometry, Ann. Otol. Rhinol. Laryngol. 85, Supplement no. 28.

Dinse, H.R. and Schreiner, C.E. (1996) Dynamic frequency tuning of the cat auditory cortical neurons: Specific adaptations to the processing of complex sounds? In: Ainsworth, W.A. and Greenberg, S. (eds.) Auditory Basis of Speech Perception, ESCA, Keele University, pp. 45-48.

Dudley, H. (1939) Remaking apeech. J. Acoust. Soc. Am. 11, 169-177.

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: Sachs, M.B. (ed.), The Physiology of the Auditory System, National Educational Consultants, Baltimore, pp. 133-141.

Greenberg, S. (1980) Temporal Neural Coding of Pitch and Vowel Quality, UCLA Working Papers in Phonetics, Vol. 52 (Ph.D. Thesis, UCLA).

Greenberg, S. (1997) The significance of the cochlear traveling wave for theories of frequency analysis and pitch. In: Lewis, E.R., Steele, C. and Lyon, R. (eds.) Diversity in Auditory Mechanics, World Scientific, Singapore.

Greenwood, D. D. (1961) Critical bandwidth and the frequency coordinates of the basilar membrane. J. Acoust. Soc. Am. 33, 1344-1356.

Jenison, R., Greenberg, S., Kluender, K. and Rhode, W. S. (1991) A composite model of the auditory periphery for the processing of speech based on the filter response functions of single auditory-nerve fibers. J. Acoust. Soc. Am. 90, 773-786.

Kiang, N.Y-S., Watanabe, T., Thomas, E.C. and Clark, L.F. (1965) Discharge Patterns of Single Fibers in the Cat's Auditory Nerve, MIT Press, Cambridge, MA.

Kraus, N. and McGee, T. (1992) Electrophysiology of the human auditory sytem, In: The Mammalian Auditory Pathway: Neurophysiology, Popper, A.N. and Fay, R. R. (eds.), Springer, New York, pp. 335-403.

Langner G. and Schreiner C.E. (1988) Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms. J. Neurophysiol.. 60, 1799-1822.

Langner, G., Schulze, H., Sams, M. and Heil, P. (1996) The topographic representation of periodicity pitch in the auditory cortex. In: Ainsworth, W.A. and Greenberg, S. (eds.) Auditory Basis of Speech Perception, ESCA, Keele University, pp. 91-97.

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.

Ragot, R. and Lepaul-Ercole, R. (1996) Brain potentials as objective indexes of auditory pitch extraction from harmonics, NeuroReport 7, 905-909.

Roberts, T. and Poeppel, D. (1996) Latency of auditory evoked M100 as a function of tone frequency, NeuroReport 7, 1138-1140.

Ruggero, M. and Rich, N. (1987) Timing of spike initiation in cochlear afferents: dependence on site of innervation, J. Neurophysiol. 58, 379-403.

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.

Schreiner, C.E., Wong, S. and Bonham, B, (1997) Spatial-temporal representation of syllables in cat primary auditory cortex, this volume.

Slaney, M. and Lyon, R. (1993) On the importance of time - a temporal representation of sound. In: Cooke, M., Beet, S. and Crawford, M. (eds.) Visual Representations of Speech Signals, Wiley, Chicester, pp. 95-118.

Smith, J.C., Marsh, J.T. and Brown, W.S. (1975) Far-field recorded frequency-following responses: Evidence for the locus of brainstem sources, Electroenceph. clin. Neurophysiol. 39, 465-472.

Winer, J. (1992) The functional architecture of the medial geniculate body and the primary auditory cortex. In: The Mammalian Auditory Pathway: Neuroanatomy, Popper, A.N. and Fay, R.R. (eds.), Springer, New York, pp. 222-409.