Neural representation of pitch through temporal autocorrelation

Peter Cariani[1,2], Mark Tramo[1,3], and Bertrand Delgutte[1,2]

[1]Eaton Peabody Laboratory, Massachusetts Eye & Ear Infirmary
[2]Department of Otology and Laryngology, Harvard Medical School
[3]Department of Neurobiology, Harvard Medical School
243 Charles St., Boston MA 02114 USA
tel (617) 573-4243, FAX (617) 720-4408
email peter@epl.meei.harvard.edu

Abstract

An enormous wealth of acoustic information is present in the temporal firing patterns of auditory neurons. Distributions of interspike intervals across neural populations in the auditory nerve and brainstem form autocorrelation-like stimulus representations that closely predict the low pitches of complex tones. Many diverse aspects of auditory perception are readily explained in terms of central analyses of these interval-based representations. To the extent that neural discharges are stimulus-locked in a given sensory system, distributions of all-order interspike intervals provide a neural representation of the stimulus autocorrelation function. These time-domain representations provide an alternative means for the nervous system to perform Fourier analysis.

The neural coding problem

The neural coding problem – how populations of neurons represent and convey information through trains of spikes – is fundamental to our understanding how sensory systems function [1,2,29]. Although a great deal is known about neural response properties at many levels of the auditory system, we presently have only a very rudimentary understanding of how auditory forms are actually represented by the central auditory system.

There are fundamentally two basic ideas about how sensory information can be encoded in patterns of neural discharge: coding by spatial patterns of neural excitation vs. coding by temporal patterns in spike trains. These alternative coding strategies could be called, respectively, "coding-by-channel" and "coding-by-time structure." Place-based or "labeled line" codes depend upon which particular neurons respond (i.e. which channels are activated). Strategies for pattern-recognition based on spatially-organized sensory maps and on specifically-tuned "feature detectors" both stem from this basic idea of coding by channel. Temporal codes, on the other hand, depend upon how neurons respond: the form of their response, rather than through which neural channels the message arrives, carries the message. Temporal codes depend upon either temporal patterns between spikes in a spike train or on the time-of-arrival of spikes relative to some reference event. Historically, coding-by-channel ideas developed from Mueller's "specific nerve energies" and Helmholtz's later resonance-place theory of auditory representation. Temporal coding ideas, on the other hand, were articulated through Rutherford's "telephone" theory of hearing, Troland's temporal-modulation representations for pitch and color, Wever's volley principle, and Licklider's duplex theory of hearing [1,22,23,40]. While the channel-coding idea has given rise to the highly developed connectionist networks of today, a theory of adaptive timing networks based on temporally-coded signals remains to be elaborated.

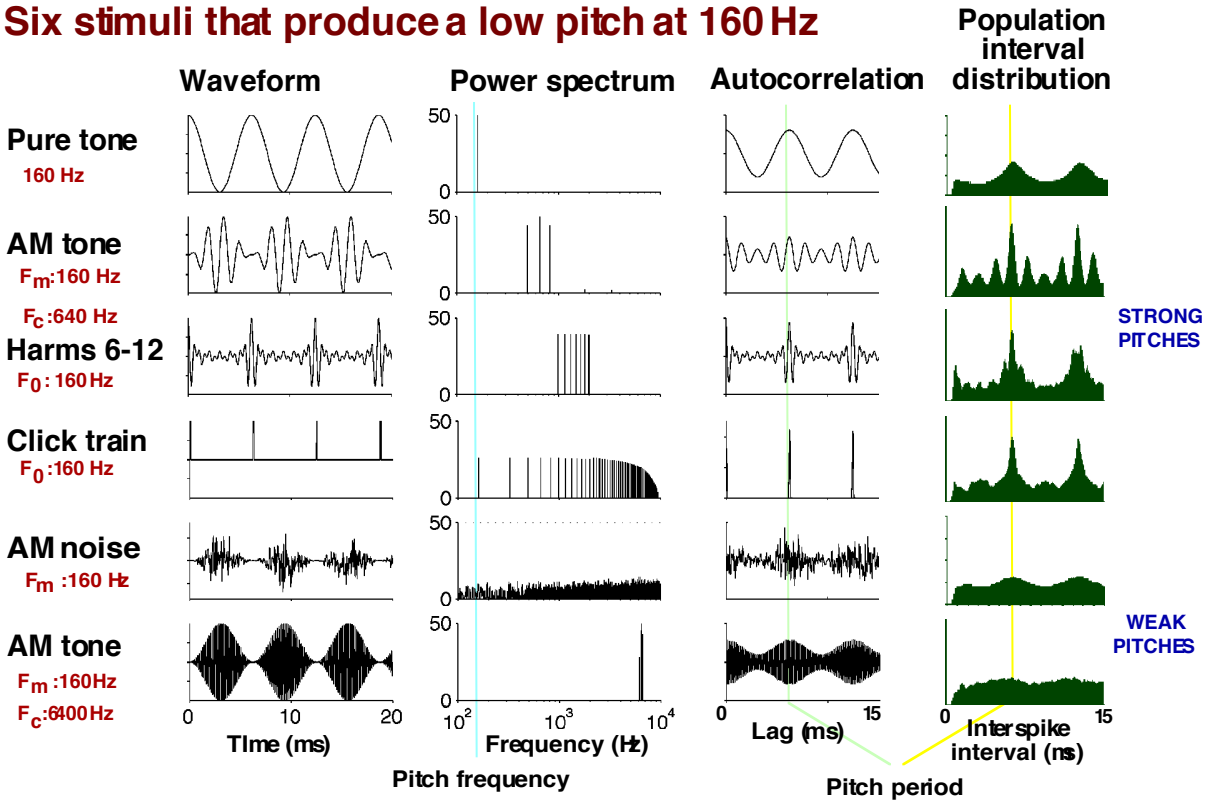

Figure 1. Neural responses to four stimuli evoking a pitch at 160 Hz but differing in pitch salience. Left to right: Stimulus waveform, power spectrum, short term autocorrelation function, and population-interval distribution for each stimulus. Population-interval distributions are constructed by summing together the all-order interspike interval distributions of many auditory nerve fibers having a wide range of characteristic frequencies (A-F: 85, 54, 49, 56, 53, 50 fibers). Histograms have been normalized to the mean number of counts/bin. Arrows indicate the position of the fundamental period (6.25 ms). A. Pure tone, 160 Hz. B. AM tone with a low frequency carrier (Fc=640 Hz, Fm=160 Hz). C. Harmonic complex (harmonics 6-12, of 160 Hz). D. Unipolar click train (F0=160 Hz.) E. AM tone with a high frequency carrier (Fc=6400 Hz, Fm=160 Hz). F. AM broadband noise with Fm=160 Hz. All stimuli presented at 60 dB total SPL.

Temporal coding of periodicity pitch

Pitch has played a pivotal role in many of the general debates about neural coding [1,12]. The mechanisms underlying the low pitches of complex tones ("periodicity pitches") have been discussed and debated for over 150 years. Throughout this history auditory physiologists and theoreticians alike have simultaneously appreciated the great abundance of information about stimulus periodicities that temporal discharge patterns of auditory neurons carry, as well as the orderly, spatial organization of the cochlea by frequency. While this general, channel-based, "place principle" has dominated thinking about neural coding in most other sensory modalities, in audition there has always been a strong case for temporal coding of pitch. The pendulum of scientific opinion has swung back and forth between spectral pattern and temporal theories [12]. Temporal autocorrelation models for pitch held sway in the 1950's [22,23], but with the discovery of the "dominance frequency region for pitch" in the early 1960's, spectral pattern models regained support. Over the last decade the temporal autocorrelation models have been revived and extended [24,25,26,35,39].

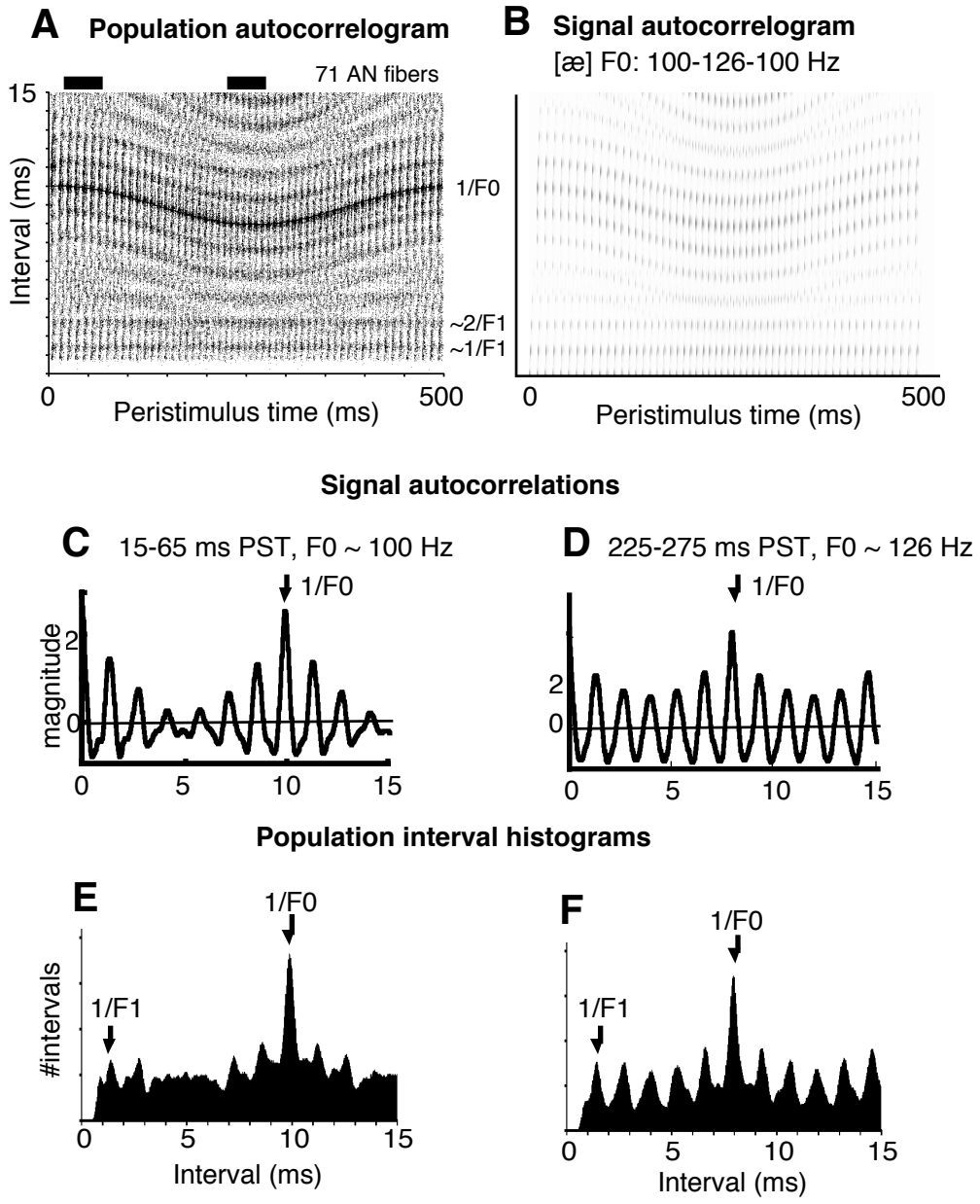

Figure 2. Autocorrelation-like representations of the vowel [æ]. Stimulus: Five-formant synthetic vowel [æ], as in "had". Formants are at 750, 1450, 2450, 3350, 3850 Hz, and the fundamental frequency F0 (heard as the voice pitch) sinusoidally varies between 100-126 Hz. A. Neural response. Running population-interval distribution (autocorrelogram) of 71 auditory nerve fibers of widely distributed characteristic frequencies. The stimulus was presented to each fiber 100 times at 60 dB SPL and running all-order interval distributions were weighted according to CF and summed. The result is an estimate of the distribution of all-order intervals across the entire auditory nerve. Each dot represents the occurence of 10 or more intervals of a given length (y ms, range 0-15 ms) ending at a given peristimulus time (x ms, range 0-500 ms). Thin line indicates the fundamental period, 1/F0, the voice pitch that would be heard, as a function of peristimulus time. B. Stimulus autocorrelogram, $SAC(\tau, t) = S(t)S(t-\tau)$, computed at 10 kHz sampling rate and thresholded. C, D. Short-time autocorrelation functions for two stimulus segments indicated in A (bars). The highest peak at the fundamental period 1/F0 corresponds to the voice pitch. E,F. Population interval histograms for the same segments.

In physiological studies at the level of the auditory nerve [7,8], we have found robust and pervasive correspondences between interspike interval statistics of populations of auditory nerve fibers and the pitches produced by a wide array of complex tones. In these studies we recorded the responses of many single auditory nerve fibers in Dial-anesthetized cats to complex stimuli that produce low, periodicity pitches in humans. We compiled all-order interspike interval distributions (i.e. counting intervals between both successive and nonsuccessive spikes) for each fiber. We weighted and summed the intervals from fibers according to their characteristic frequencies (CFs) in order to estimate what the all-order interval distribution would be for the entire population of auditory nerve fibers in a human listener. The result is a population-interval distribution, the distribution of intervals thought to be present in the entire auditory nerve. These kinds of interval-based representations constitute a possible means by which the auditory system might represent the structure of sounds. Such temporal neural representations complement those channel-based representations that are based on spatial patterns of discharge in auditory frequency maps.

With very few exceptions, we found that the most common all-order interval present in the population corresponds to the pitch that is heard. This can be seen in Figure 1, which shows the waveforms, power spectra, short-term autocorrelation functions for five stimuli that produce definite pitches at 160 Hz. A sixth stimulus (D) lies just outside the classical existence region for periodicity pitch, and produces a very weak, ill-defined pitch. Several of these stimuli (B, C, E) have "missing fundamentals" at 160 Hz. The population-interval distributions for these stimuli at the level of the auditory nerve are shown in the rightmost panels. In all cases, the positions of major interval peaks correspond to the period of the pitch that is heard (i.e. the fundamental period for harmonic complexes or the modulation period for AM noise) and its multiples. We found this rule to hold at low (40 dB SPL), moderate (60 dB) and high (80 dB) stimulus levels, and at all signal-to-noise ratios where the pitch could be heard. This suggests that all-order interval codes provide extremely robust representations of pitch, that, like the pitch percept itself, are not greatly distorted or degraded by high levels or background noise.

Our second major finding was that the relative proportion of pitch-related intervals amongst all intervals qualitatively corresponded to pitch strength. In Figure 1, stimuli A-D evoke strong periodicity pitches, whereas stimuli E and F evoke much weaker pitches. Correspondingly, in their respective population-interval distributions, the peak-to-background ratios of the major, pitch-related interval peaks are much higher for those stimuli (A-D) that produce strong periodicity pitches.

These findings taken together with the rest of our data suggest that many diverse aspects of pitch can be directly explained in terms of population-interval distributions at the level of the auditory nerve: the pitch of the "missing fundamental", pitch equivalence of stimuli with very different power spectra, pitch shifts and pitch ambiguities produced by inharmonic AM tones, the relative phase- and level-invariance of periodicity pitches, pitches produced by unresolved harmonics and by AM noise, and the dominance (frequency) region for pitch. From studies of neural responses in the auditory brainstem [3,4,19,31], it appears that population-interval distributions can serve as representations for periodicity pitch in the central auditory system, although the extent to which pitch-related timing information exists at the level of primary auditory cortex is still unclear.

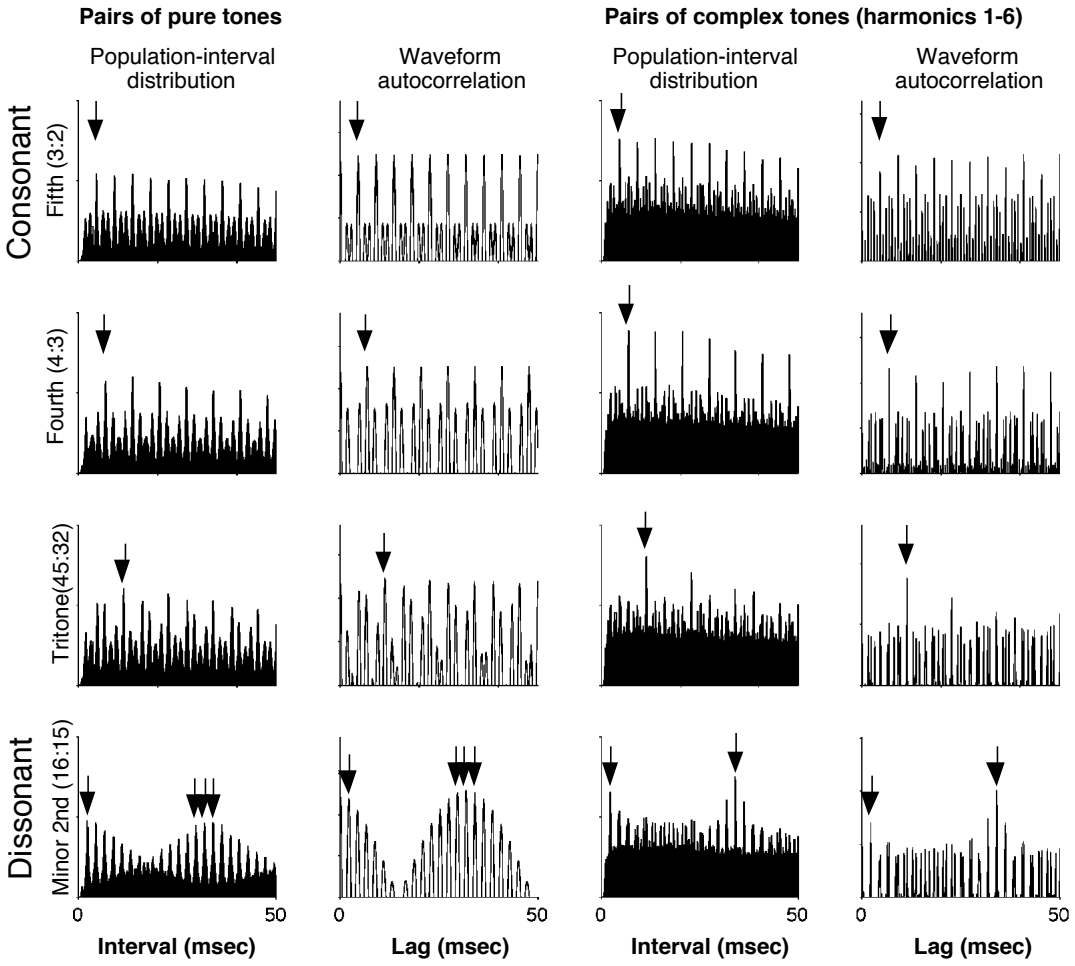

Figure 3. Population-interval neural representations and corresponding stimulus autocorrelation functions for five musical intervals. Stimuli were pairs of pure tones or pairs of complex tones consisting of harmonics 1-6 (equal amplitudes), with the lower fundamental frequency = 440 Hz. For complex tones, the fifth (3:2) and the fourth (4:3) are generally regarded as consonant intervals with the tritone (45:32) and minor second (16:15) being dissonant ones. For pure tones, the rank ordering follows frequency separation, with the tritone being slightly more consonant than the fourth. All 200 msec stimuli were presented at 60 dB SPL, 100 times to each auditory nerve fiber (ANF) through a closed, calibrated acoustic assembly. Population-interval distributions were compiled by summing together the all-order interspike interval distributions of 50-100 ANF's distributed across a wide range of characteristic frequencies (CF's). Positive portions of stimulus autocorrelation functions are shown. Arrows indicate positions of major peaks in population-interval and autocorrelation plots that may correspond to low pitches.

Temporal autocorrelation

In our investigations of the representation of pitch at the level of the auditory nerve we also observed that that the forms of population-interval distributions resembled their respective stimulus autocorrelation functions. In this paper we will discuss this observation more deeply, outlining in what sense these population-interval distributions might constitute autocorrelation-like representations of the stimulus. We will briefly explore what this might mean in terms of auditory representations and neural information processing.

Correlational analysis [20,41] was a much more widely used for the analysis of complex signals in the decades before the "rediscovery" of the fast Fourier transform in the 1960's and the general availability of digital signal processing. Delay lines were common in the computing machinery of the day, and analog autocorrelators existed for real-time analysis of time-series inputs [20]. Temporal autocorrelation and cross-correlation architectures were proposed for the representation and the separation of auditory forms [9,23,36]. In other sensory modalities, temporal correlation mechanisms successfully explained many aspects of motion detection in the fly visual system [30] and flutter-vibration frequency discrimination in the somatosensory system [27]. Interspike intervals and temporal discharge patterns have also been proposed as means of multiplexing different kinds of visual information in neural pathways [10,32].

Temporal correlation functions entail the multiplication of a time-series signal by another signal at different relative time delays (τ). Cross-correlation entails the multiplication of two different signals, while autocorrelation entails the multiplication of a signal by itself. Because the output of the multiplication operation is a joint property of both signals, correlation functions reflect patterns of joint occurrences of events in time.

Correlation functions are intimately related to Fourier transforms. Temporal correlations are expressed as functions of time delays (time domain), while Fourier transforms and power spectra are expressed as functions of frequency (frequency domain). Power spectra can be obtained by computing the Fourier transform of the autocorrelation function. Because this operation is reversible (invertible), the autocorrelation function of a signal contains exactly the same information about a stimulus as its power spectrum. Compared to the original waveform, the autocorrelation function and the power spectrum retain periodicity information while discarding phase information.

Currently the power spectrum and the spectrogram are most commonly used representations of sound. Correspondingly, the auditory system is most often conceptualized of in terms of spatially-distributed spectral representations, where particular sub-populations of auditory neurons are tuned to particular frequency ranges. The profile of average discharge rates across these tonotopic auditory frequency maps in effect provides a neural representation of the stimulus power spectrum. These rate-place representations are channel-based: in their simplest forms, each neuron represents the frequency to which it is maximally tuned, the discharge rate encoding in some fashion the amount of stimulus energy in that frequency neighborhood.

Neural time domain representations of the stimulus waveform consist of times of arrivals in the spike trains themselves. Time domain representations of the stimulus autocorrelation function consist of distributions of interspike intervals, the times between spikes. The autocorrelation function is simply a time-series multiplied by itself as a function of relative time delay: $f(\tau) = \Sigma\, S(t)\, S(t-\tau)$, summed over all times t. Typically we compute the (retrospective) autocorrelation function only for short, positive time lags that are in the periodicity pitch range (e.g. 0-15 ms), thereby reducing the amount of computation required. Autocorrelation functions are intimately related to interspike interval histograms. If a spike train is described in terms of a time-series of 0's (no spike

in a given time-bin) and 1's (spike in a given time-bin), then the retrospective autocorrelation function of the spike train is the same as a histogram of the time intervals between all possible pairs of spikes in the train, i.e. between both successive and non-successive spikes.

Temporal autocorrelation and all-order interspike intervals

We found through observation that the delay-positions of major and minor peaks in population-interval distributions closely mirror those of the stimulus autocorrelation function. This can be readily appreciated by comparing in Figure 1 the population-interval distributions with their respective short-time stimulus autocorrelation functions. This correspondence holds to the extent that there is phase-locking, i.e. for periodicities up to a few kHz. While all of the major and minor peaks in autocorrelation of the AM tone with the low-frequency carrier (B) are replicated in its population-interval response, the corresponding fine structure of the AM tone with the high-frequency carrier (E) is completely absent: only a very shallow remnant of the envelope remains.

In another series of experiments, we observed the responses of auditory nerve fibers to five synthetic vowels [5,6]. These vowels had fundamental frequencies (F0s) that varied sinusoidally between 100-126 Hz. Stimulus autocorrelation functions and population-interval distributions for the vowel [ae] are shown in Figure 2. The population autocorrelogram, or running population-interval distribution, and the signal autocorrelogram are shown in panels A and B. The voice pitch of such a vowel is heard at the fundamental period. The densest interval band in the population autocorrelogram (A) closely followed the fundamental period and voice pitch period throughout the entire range of fundamental frequencies presented. There is strong correspondence between the structure of the population and signal autocorrelograms in panels A and B.

In panels C-F the signal short-time autocorrelation functions and population interval histograms are shown for two vowel segments, when F0 = 100 Hz and when F0 = 126 Hz. Major peaks at 10 ms and at 8 ms correspond to respectively to the fundamental periods (and voice pitches) or the two segments. Each vowel has a characteristic autocorrelation function, and we found that in general the population-interval distributions resembled their respective stimulus autocorrelation functions. Essentially each vowel's characteristic format structure sets up characteristic autocorrelation and interval patterns. This is consistent with previous physiological observations from both single ANFs and ensembles of ANFs [14,28]. We found that vowels could be discriminated on the basis of population-interval distributions consisting of short intervals (0-5 ms), i.e. on the basis of temporal information alone. This suggests that the timbres of stationary sounds with low- and medium-frequency components may be explicable in purely temporal terms [24]. In other experiments we have found that changes in population-interval distributions covary with vowel-class boundaries [18].

We have also used stimuli that have varying qualities of musical consonance or dissonance [38]. These stimuli consisted of either pairs of pure tones or pairs of complex tones (harmonics 1-6). The (fundamental) frequency of the first tone was always 440 Hz. The second frequency was separated from the first by various musical intervals (frequency ratios): fifth (3:2), fourth (4:3), tritone (45:32), or minor second (16:15). The population-interval distributions and the positive portions of their respective stimulus autocorrelation functions are shown in Figure 3. We are currently analyzing the neural responses to these stimuli using a variety of representations, including population-interval distributions, to explore possible neural correlates of musical consonance, roughness, fusion, and distributions of perceived pitch(es). In music theory, which primarily deals with relations between the complex tones produced by musical instruments, the musical fifth and fourth are generally considered to be consonant intervals, whereas the tritone and

the minor second are generally considered as dissonant. Qualitatively, the more consonant stimuli produce population-interval distributions with simpler, repetitive interval patterns, whereas the dissonant stimuli produce more complex, less repetitive ones. What is most striking in this context, however, is the similarity between the population-interval distributions and their respective stimulus autocorrelation functions. For virtually all sets of stimulus autocorrelation peaks there exist corresponding sets of interval peaks. The one exception is for the pure tone fourth (4:3), where there is an extra set of small peaks at half the fundamental period and its multiples (this corresponds to the distortion product $2f_1 - f_2$, which, for the 4:3 ratio of the fourth, equals 2F0).

In retrospect, the reasons that population-interval distributions should resemble stimulus autocorrelation functions are fairly straightforward. They depend mainly on the phase-locked nature of auditory nerve fiber discharges. Each stimulus component produces discharges that are phase-locked discharges to itself, predominantly, but not limited to, those auditory nerve fibers whose characteristic frequencies (CF) are nearby. In doing so, intervals at subharmonics, integer multiples of the component's period, are produced. If the stimulus is a harmonic complex, then all stimulus components have a common subharmonic at the fundamental. When all of the intervals corresponding to all of the subharmonics are summed together in a population interval distribution, the most common intervals are invariably at the fundamental period and its multiples (i.e. the fundamental frequency and its subharmonics). This is the time-domain equivalent of Terhardt's frequency-based method of subharmonic coincidence [37]. If interspike intervals are the means by which the auditory system represents pitch, then central auditory analyzers interpret the interval-pattern associated with the fundamental frequency (even if it is "missing") as a low pitch. Thus, the perception of periodicity pitch could well be a direct consequence of the basic neural codes that the auditory system uses coupled with the phase-locked, stimulus-driven character of its neural discharges. If this is the case, pitch judgments are well-described by temporal autocorrelation models [11,12] precisely because the neural representations that subserve those judgments are themselves autocorrelation-like.

The similarity extends beyond the patterns of major peaks that are associated with periodicity pitch. Minor interval peaks are produced by other combinations of intervals (subharmonics of stimulus components). These patterns repeat at each fundamental period in the autocorrelation function (note the repeating patterns for the consonant stimuli and responses of Figure 3 and in the autocorrelations of Figure 2 in panels D and F). Different vowels with different formant regions (and different timbres) give rise to different repeating patterns of minor peaks that are nested within the F0-related major peaks. It is not surprising then, that the delay-positions of interval peaks should mirror those of the stimulus, since each stimulus component produces intervals related to it.

In some cases (Figure 3), the relative heights of peaks are similar in stimulus autocorrelations and population-interval distributions, whereas in others (Figs. 1B and 2), the relative heights are noticeably different. These similarities and differences may stem from the degree to which cochlear filtering and spike initiation are linear processes. Autocorrelations of individual frequency components summed together equal the autocorrelation of the whole [22]. For an array of contiguous band-pass linear filters of uniform bandwidth, the sum of the channel autocorrelations is proportional to the autocorrelation of the unfiltered stimulus. To the extent that the production of intervals is the result of a linear process, then summing the intervals should yield relative amplitudes that mirror the stimulus autocorrelation function. To the extent that nonlinearities are created by cochlear filtering, hair cell transduction (half-wave rectification), synaptic transmission, nonuniform distributions of characteristic frequencies of auditory nerve fibers or their rate-level

functions (threshold and saturation effects), the respective heights of interval peaks will diverge from their counterparts in the stimulus autocorrelation function.

The functional effects of nonlinearities in the auditory system depend critically on the nature of the neural representations, i.e. what aspects of the neural signal are actually used by the auditory system to subserve a given auditory percept. Interval-based representations of stimulus periodicities appear to be relatively resilient to the introduction of many of the above-mentioned nonlinearities. Nonlinear changes in discharge rates with level do not distort the time intervals that correspond to a particular frequency component – they merely cause relatively fewer or more intervals associated with that component to be produced. Combination tones created by nonlinear distortion produce sets of related intervals that either augment those associated with stimulus components or add entirely new sets of intervals to the distribution [16], as was seen for the pure tone fourth in Figure 3. In population-interval distributions the delay positions of the interval peaks themselves are generally unchanged, only the relative heights of peaks are altered. Thus, under a population-interval code, information about the frequency of a stimulus component is generally not degraded, while the information concerning the relative intensity of that component may be considerably distorted by nonlinearities. Population-interval distributions therefore appear to be more faithful in their representation of the frequencies of stimulus components that are present than they are in their representation of the relative intensities of those components. In many ways this behavior parallels our auditory perceptions. Subtle changes in the relative levels of stimulus components generally affect the quality of sounds far less than comparable changes in component frequencies. For example, for pure tones at 1 kHz and moderate levels, the difference limens for intensity, expressed in terms of Weber fractions $(I + \Delta I)/I$, are some 40 times those for frequency, $(f + \Delta f)/f$ [33]. Interestingly, given the discharge properties of auditory nerve fibers, autocorrelation analysis is by far the decision strategy that most closely approaches the performance of the ideal pure tone frequency discriminator [13,17,34]. Independent of whether the central auditory system utilizes such temporal coding strategies to represent auditory forms, receptor arrays capable of phase-locking coupled with temporal autocorrelation analysis offer extremely powerful and robust strategies for discriminating sounds, strategies that we have only barely begun to incorporate into devices for processing audio signals [15,21,24,35].

Conclusions

The potential implications of autocorrelation-like representations in the auditory system are many. Neural codes based on interspike intervals allow the stimulus power spectrum to be represented and analyzed in the time domain. To the extent that there is phase-locking of neural discharges to stimulus components, such interval-based codes can form the basis of stimulus representations that complement spectral, tonotopically-based rate-place ones. Periodicity pitch along with many other aspects of auditory perception may thus be direct consequences of the kinds of temporally-based neural representations that the auditory system employs for the analysis of sounds. Major questions remain for how such temporal information might be utilized by the central auditory system to give rise to some of the qualities of sound that we hear. We need to better understand the extent to which the timing information that we observe in the auditory nerve is available at higher auditory centers as well as the neural computational strategies by which this information might be effectively used.

Acknowledgments

This work was supported by Grant DC03054 from the National Institute for Deafness and Communications Disorders (NIDCD), of the National Institutes of Health (NIH).

References

1. Boring, E.G. 1942, *Sensation and Perception in the History of Experimental Psychology,* Appleton-Century-Crofts, New York.
2. Cariani, P., 1995, As if time really mattered: temporal strategies for neural coding of sensory information, *Communication and Cognition - Artificial Intelligence (CC-AI)* 12(1-2): 161-229. Reprinted in: Origins: Brain and Self-Organization, (K. Pribram, ed.), Lawrence Erlbaum, Hillsdale, NJ, 1995.
3. Cariani, P., 1995, Physiological correlates of periodicity pitch in the cochlear nucleus, *Assoc. Res. Otolaryn. Abstr.* : 128.
4. Cariani, P., 1997, Population-interval representations of pitch in the auditory brainstem, *Assoc. Res. Otolaryn. Abstr.* :
5. Cariani, P. and B. Delgutte, 1993, Interspike interval distributions of auditory nerve fibers in response to concurrent vowels with same and different fundamental frequencies, *Assoc. Res. Otolaryngology. Abs.* : 373.
6. Cariani, P. and B. Delgutte, 1994, Transient changes in neural discharge patterns may enhance separation of concurrent vowels with different fundamental frequencies [Abstract], *J. Acoust. Soc. Am.* 95(5(2)): 2842.
7. Cariani, P.A. and B. Delgutte, 1996, Neural correlates of the pitch of complex tones. I. Pitch and pitch salience., *J. Neurophysiol.* 76(3): 1698-1716.
8. Cariani, P.A. and B. Delgutte, 1996, Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase-invariance, pitch circularity, and the dominance region for pitch, *J. Neurophysiol.* 76(3): 1717-1734.
9. Cherry, C., 1961, Two ears – but one world, in: *Sensory Communication,* (W. A. Rosenblith, ed.), MIT Press/John Wiley, New York.
10. Chung, S.H., S.A. Raymond and J.Y. Lettvin, 1970, Multiple meaning in single visual units, *Brain Behav Evol* 3: 72-101.
11. de Boer, E. *On the "residue" in hearing.* Ph.D., University of Amsterdam (1956)
12. de Boer, E., 1976, On the "residue" and auditory pitch perception, in: *Handbook of Sensory Physiology,* (W. D. Keidel and W. D. Neff, ed.), Springer Verlag, Berlin.
13. Delgutte, B., 1995, Physiological models for basic auditory percepts, in: *Auditory Computation,* (H. Hawkins, T. McMullin, A. N. Popper and R. R. Fay, ed.), Springer Verlag, New York.
14. Delgutte, B. and N.Y.S. Kiang, 1984, Speech coding in the auditory nerve: I. Vowel-like sounds, *J. Acoust. Soc. Am.* 75(3): 866-878.
15. Ghitza, O., 1992, Auditory nerve representation as a basis for speech processing, in: *Advances in Speech Signal Processing,* (S. Furui and M. M. Sondhi, ed.), Marcel Dekker, New York.
16. Goldstein, J.L. and N.Y.S. Kiang, 1968, Neural correlates of the aural combination tone $2f_1$-f_2, *IEEE Proc* 56: 981-992.
17. Goldstein, J.L. and P. Srulovicz, 1977, Auditory-nerve spike intervals as an adequate basis for aural frequency measurement, in: *Psychophysics and Physiology of Hearing,* (E. F. Evans and J. P. Wilson, ed.), Academic Press, London.
18. Hirahara, T., P. Cariani and B. Delgutte, 1996, Representation of low-frequency vowel formants in the auditory nerve, in: *Proceedings, ESCA Research Workshop on The Auditory Basis of Speech Perception, Keele University, United Kingdom, July 15 - 19, 1996,* ed.),
19. Kim, D.O. and G. Leonard, 1988, Pitch-period following response of cat cochlear nucleus neurons to speech sounds, in: *Basic Issues in Hearing,* (H. Duifhuis, J. W. Horst and H. P. Wit, ed.), Academic Press, London.
20. Lange, F.H. 1967, *Correlation Techniques,* Van Nostrand, Princeton.
21. Lazzaro, J. and C. Mead, 1989, Silicon modeling of pitch perception, *Proc. Nat. Acad. Sci. U.S.A.* 86: 9597-9601.
22. Licklider, J.C.R., 1951, A duplex theory of pitch perception, *Experientia* VII(4): 128-134.
23. Licklider, J.C.R., 1959, Three auditory theories, in: *Psychology: A Study of a Science. Study I. Conceptual and Systematic,* (S. Koch, ed.), McGraw-Hill, New York.

24. Lyon, R. and S. Shamma, 1995, Auditory representations of timbre and pitch, in: *Auditory Computation,* (H. Hawkins, T. McMullin, A. N. Popper and R. R. Fay, ed.), Springer Verlag, New York.
25. Meddis, R. and M.J. Hewitt, 1991, Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I. Pitch identification, *J. Acoust. Soc. Am.* 89(6): 2866-2882.
26. Moore, B.C.J. 1989, *Introduction to the Psychology of Hearing,* Academic Press, London.
27. Mountcastle, V., 1993, Temporal order determinants in a somatosthetic frequency discrimination: sequential order coding, *Annals New York Acad. Sci.* 682: 151-170.
28. Palmer, A.R., 1988, The representation of concurrent vowels in the temporal discharge patterns of auditory nerve fibers, in: *Basic Issues in Hearing,* (H. Duifhuis, J. W. Horst and H. P. Wit, ed.), Academic Press, London.
29. Perkell, D.H. and T.H. Bullock, 1968, Neural Coding, *Neurosciences Research Program Bulletin* 6(3): 221-348.
30. Reichardt, W., 1961, Autocorrelation, a principle for the evaluation of sensory information by the central nervous system, in: *Sensory Communication,* (W. A. Rosenblith, ed.), MIT Press/John Wiley, New York.
31. Rhode, W.S., 1995, Interspike intervals as correlates of periodicity pitch in cat cochlear nucleus, *J. Acoust. Soc. Am.* 97(4): 2414-2429.
32. Richmond, B.J., L.M. Optican and T.J. Gawne, 1989, Neurons use multiple messages encoded in temporally modulated spike trains to represent pictures, in: *Seeing Contour and Colour,* (J. J. Kulikowski and C. M. Dickenson, ed.), Pergamon Press, New York.
33. Siebert, W.M., 1965, Some implications of the stochastic behavior of primary auditory neurons, *Kybernetik* 2: 206-215.
34. Siebert, W.M., 1970, Frequency discrimination in the auditory system: place or periodicity mechanisms?, *Proc IEEE* 58: 723-730.
35. Slaney, M. and R.F. Lyon, 1993, On the importance of time - a temporal representation of sound, in: *Visual Representations of Speech Signals,* (M. Cooke, S. Beet and M. Crawford, ed.), John Wiley, New York.
36. Stevens, K.N., 1950, Autocorrelation analysis of speech sounds, *J. Acoust. Soc. Am.* 22(6):769-71.
37. Terhardt, E., G. Stoll and M. Seewann, 1982, Algorithm for extraction of pitch and pitch salience from complex tonal signals, *J. Acoust. Soc. Am.* 71(3): 679-688.
38. Tramo, M.J., P. Cariani and B. Delgutte, 1992, Representation of tonal consonance and dissonance in the temporal firing patterns of auditory nerve fibers, *Soc. Neurosci. Abstr.* 18: 382.
39. van Noorden, L., 1982, Two channel pitch perception, in: *Music, Mind and Brain,* (M. Clynes, ed.), Plenum, New York.
40. Wever, E.G. 1949, *Theory of Hearing,* Wiley, New York.
41. Wiener, N. 1961, *Cybernetics: or Control and Communication in the Animal and in the Machine,* MIT Press, Cambridge, MA.

Hirahara T, Cariani P, and Delgutte B. Representation of low-frequency vowel formants in the auditory nerve. In: Proceedings, ESCA Research Workshop on The Auditory Basis of Speech Perception, Keele University, United Kingdom, July 15 - 19, 1996.

REPRESENTATION OF LOW-FREQUENCY VOWEL FORMANTS IN THE AUDITORY NERVE

Tatsuya Hirahara[1], Peter Cariani[2], and Bertrand Delgutte[2,3]
hirahara@idea.brl.ntt.jp peter@epl.meei.harvard.edu bard@epl.meei.harvard.edu

[1] NTT Basic Research Laboratories
3-1 Morinosato-Wakamiya
Astugi, Kanagawa 243-01
JAPAN

[2] Eaton Peabody Laboratory
Massachusetts Eye and Ear Infirmary
Boston, MA 02114
U.S.A.

[3] Research Laboratories of Electronics
Massachusetts Institute of Technology
Cambridge, MA 02139
U.S.A.

[ABSP-1996] Auditory Basis of Speech Perception, Keele, UK, July 15-19, 1996, ed. by William A. Ainsworth and Steve Greenberg, ISCA Archive, http://www.isca-speech.org/archive/absp_96

REPRESENTATION OF LOW-FREQUENCY VOWEL FORMANTS IN THE AUDITORY NERVE

Tatsuya Hirahara[1], Peter Cariani[2], and Bertrand Delgutte[2,3]
hirahara@idea.brl.ntt.jp peter@epl.meei.harvard.edu bard@epl.meei.harvard.edu

[1] NTT Basic Research Laboratories
3-1 Morinosato-Wakamiya
Astugi, Kanagawa 243-01
JAPAN

[2] Eaton Peabody Laboratory
Massachusetts Eye and Ear Infirmary
Boston, MA 02114
U.S.A.

[3] Research Laboratories of Electronics
Massachusetts Institute of Technology
Cambridge, MA 02139
U.S.A.

ABSTRACT

We have investigated the auditory representation of vowels with low-frequency formants by recording the activity of auditory-nerve fibers in anesthetized cats in response to Japanese /i/-/e/ synthetic-vowel continua. Vowels having either low (150 Hz) or high (350 Hz) fundamental frequency F0 were varied in either first-formant frequency F1 or the level of a "crucial harmonic" near F1 to span the /i/-/e/ continuum. Two different neural representations of the stimulus spectrum in the F1 region were examined: a population rate-place profile and a population interspike interval distribution. Characteristics of both representations depend on F0. When individual harmonics are resolved by the ear, as for high F0s, first formant frequency does not have explicit correlates in either ANF rate-place patterns or interspike interval distributions. Rather, both representations show clear patterns corresponding to individual harmonics, as well as the amplitude ratios of "crucial harmonics" near F1 that determine vowel identity in psychophysical tests. When harmonics are not resolved, as for low F0s, both rate-place and population-interval profiles of individual harmonics fuse to form broader, single peaks near F1, providing an explicit neural representation of formant frequency.

1. INTRODUCTION

Formant frequencies are important for vowel identification, yet the neural representation of formants is poorly understood, particularly in low frequency regions. Formants are resonant frequencies of the vocal tract which appear as local maxima (peaks) in the envelope of the stimulus spectrum. In general, spectral energy is present at harmonics of the fundamental frequency (F0) rather than at the formant frequency. For vowel discrimination, the auditory system could use a representation based on either the spectral envelope (formant) or the fine spectral structure (harmonics). Most models of speech perception assume that vowel quality is based on a single peak at the formant frequency in a smeared internal spectrum. However, low-frequency (<1000 Hz) harmonics are resolved by the human auditory system, so that psychoacoustic excitation patterns [1] exhibit separate peaks for individual harmonics rather a single formant peak.

Recent psychophysical results[2] suggest that harmonic fine structure near low-frequency formants plays an important role in vowel perception. Specifically, the phonetic boundary between the Japanese vowels /i/ and /e/ appears to be primarily determined by the amplitude ratio of two crucial harmonics that are nearest the first formant frequency (F1). Normally this amplitude ratio is determined by both F0 and F1. In these experiments, two synthetic-vowel continua spanning /i/ to /e/ were constructed. One is an F1-continuum in which the frequency of a resonator was systematically shifted, thereby altering the relative amplitudes of all harmonics near F1. Another is an L(nF0)-continuum in which the amplitude of a single crucial harmonic was systematically varied. The perceptual boundary between /i/ and /e/ was found to be the same for both kinds of stimulus manipulations when expressed in terms of an amplitude ratio of the crucial harmonics (Fig. 1). Further experiments showed that changes in F0 can influence vowel quality both by determining which harmonics are crucial and by altering the amplitude ratio at the boundary.

In the present electrophysiological study, vowel stimuli from Hirahara's psychophysical experiments[2] were used to answer two questions about how formants are represented in the auditory nerve: (1) under what conditions is formant frequency rather than fine harmonic structure explicitly represented? (2) which neural representation corresponds best to human judgments of vowel quality, a population rate-place profile or a population interspike interval distribution? We are particularly interested in high-F0 (>200 Hz) vowels whose auditory-nerve representation has not been described.

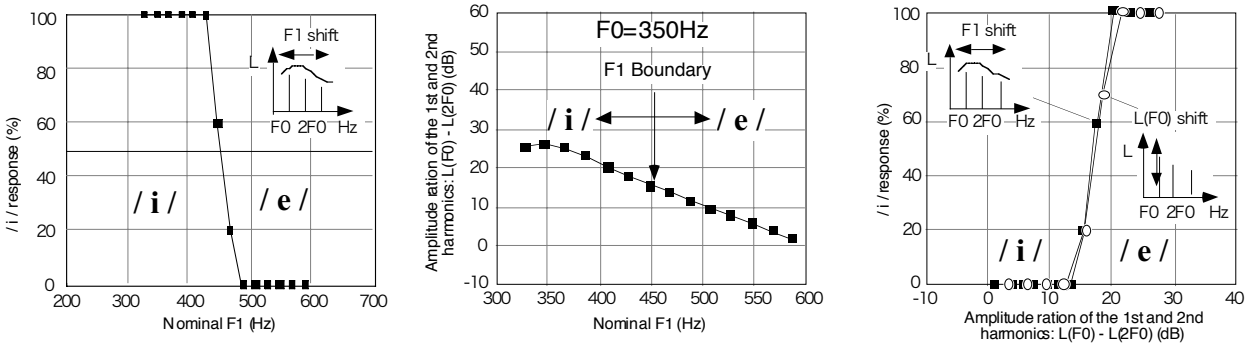

Fig. 1 Mean response curve of four subjects for the high-F0 (350Hz) F1-continuum (left). The amplitude ratio of the first and second harmonics of the stimuli covaries with F1 (middle). The perceptual vowel boundary between /i/ and /e/ are the same for the F1-continuum and the L(F0)-continuum when expressed in terms of the crutial harmonics amplitude ratio (right) [2].

2. METHODS

2.1 Stimuli

Stimuli were synthetic vowels forming continua between the Japanese vowels /i/ and /e/. All stimuli were produced by a 6-formant synthesizer using a 20 kHz sampling rate. Two types of continua were generated, each one using both a low F0 (150 Hz) and a high F0 (350 Hz). F1-continua varied the first formant frequency (F1) from 328 Hz to 528 Hz in 20-Hz steps. In these continua, amplitudes of all harmonics near F1 change as F1 increases. The phonetic boundary between /i/ and /e/ occurs roughly at F1=328 Hz for the low-F0 continuum, and at F1=448 Hz for the high-F0 continuum. Stimuli with F1 below the boundary are heard as /i/. L(nF0)-continua in which the amplitude of a single harmonic with frequency nF0 was varied in 3 dB steps over a 115 dB range that was centered at the /i/-/e/ phonetic boundary. For the low F0 continuum, the second harmonic (2F0) was manipulated, while for the high-F0 continuum, the first harmonic (F0) was manipulated. Stimuli with harmonic amplitudes above the boundary are heard as /i/.

For efficient data collection, the 11 stimuli forming each (F1- or L(nF0)) continuum were concatenated into an ascending-descending sequence. Each of the 20 stimuli in a sequence was approximately 50 msec in duration, giving a total duration of 1 sec.

2.2 Electrophysiological Recordings

The activity of auditory-nerve fibers was recorded in Dial-anesthetized cats using glass micropipettes. Stimuli were delivered through calibrated, closed acoustic assemblies at an overall level of 50 dB SPL. Times of action potentials were recorded with a precision of 1 μsec.

Once a fiber was isolated, a threshold tuning curve and spontaneous discharge rate (SR) were determined. An accurate estimate of the fiber CF was obtained with broadband-noise stimuli using de Boer's reverse correlation technique. A rate-level function for a tone at the CF was measured to determine the maximum discharge rate. Each vowel sequence was presented 60 times per fiber.

2.3 Data Analysis

For the rate-place analysis, normalized driven discharge rate was computed by averaging raw fiber discharge rate over the entire stimulus duration, subtracting out the spontaneous rate to obtain the driven rate, then normalizing the driven rate to a dimensionless quantity between 0 and 1 by dividing by the maximum driven rate. Rate-place profiles were formed by plotting normalized driven rate against fiber CF. A moving-window average of these profiles was obtained using 150-Hz windows.

For the interspike-interval analysis, an all-order interspike interval histogram was computed. These histograms include intervals between non-consecutive as well as consecutive spikes. Interspike intervals from all fibers with CFs below 1000 Hz were summed to form pooled interspike interval distributions.

These results are based on recordings from 235 auditory-nerve fibers in five cats.

3 RESULTS

3.1 Rate-place representation

Figure 2 shows the rate-place representation of the low-F0 F1-continuum. The left panel shows the normalized discharge rate as a function of both CF and F1 for all stimuli. Darker areas correspond to higher rates. Individual harmonics are clearly not resolved and maximum rates occur roughly when CF=F1. The plot resembles a smeared power spectrum of the vowel, with a broad peak near F1 [3][4].

Right panels are rate-place profiles for 3 vowel stimuli within the continuum, i.e. the horizontal cross sections of the left 2D diagram. Each data point represents normalized discharge rate for one auditory-nerve fiber. The solid line is a moving-average of the data points. For all 3 stimuli, the rate-place profile shows a maximum roughly centered at CF=F1. No peak is found at individual harmonics (300, 450, 600 Hz). Thus there is an explicit representation of F1 in rate-place profiles for low-F0 stimuli, when

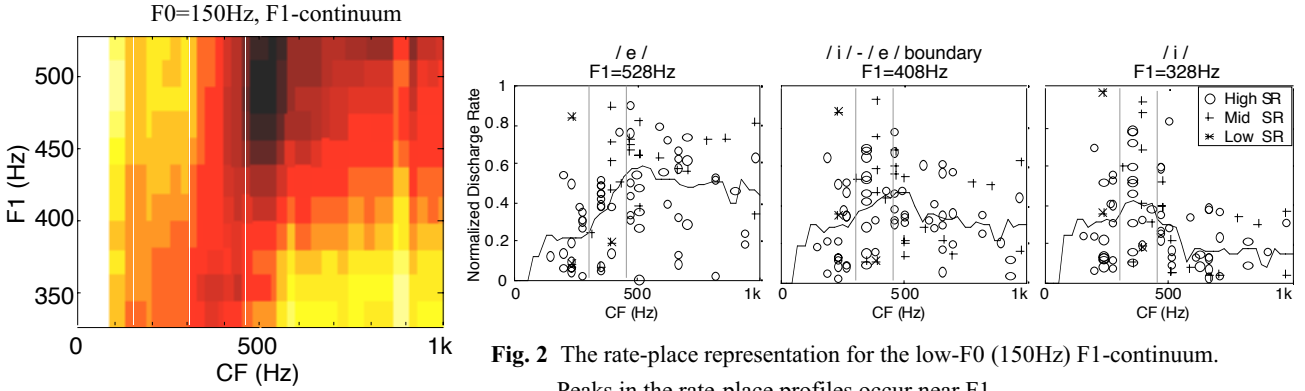

Fig. 2 The rate-place representation for the low-F0 (150Hz) F1-continuum. Peaks in the rate-place profiles occur near F1.

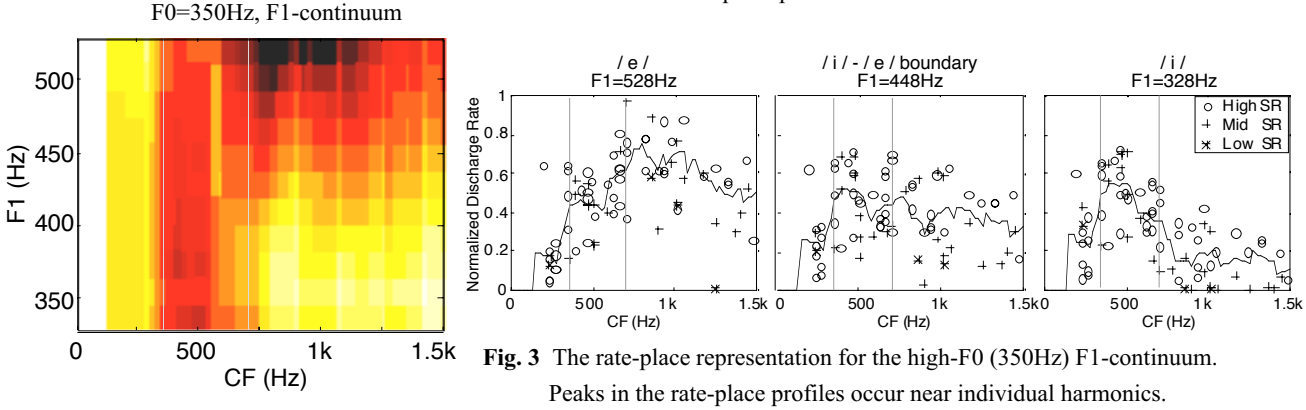

Fig. 3 The rate-place representation for the high-F0 (350Hz) F1-continuum. Peaks in the rate-place profiles occur near individual harmonics.

individual harmonics are not resolved by the ear.

Figure 3 shows the rate-place representation of the high-F0 F1-continuum. In contrast to the low-F0 continuum, population rate-place profiles for high-F0 stimuli show multiple peaks at individual harmonics rather than a single broad peak at F1. Maximum rates thus occur when CF=nF0 (here 350, 700, 1050 Hz), rather than when CF=F1. This implies that for higher-F0s, there is an explicit representation of individual, resolved harmonics rather than one for formant frequency F1.

3.2 Interspike-interval representation

Figure 4 shows the interspike-interval representation for the low-F0 F1-continuum. The left panel shows pooled interval distribution as a function of F1 for all stimuli in the continuum. Darker areas correspond to greater numbers of intervals. While the main modes are always at n/F0, secondary modes change systematically with F1. Thus, the pooled interval distribution gives an explicit representation of F1 for low F0s, when individual harmonics are not resolved in rate-place profiles.

Right panels show pooled interspike-interval distributions for 3 stimuli in the continuum, i.e. horizontal cross-sections of the left 2D diagram. For all 3 stimuli, the highest peaks in the pooled distribution are at the fundamental period 1/F0 and its multiples [5]. In addition, for both the /e/ stimulus and the /i/-/e/ boundary, secondary modes of the pooled distribution are found at 1/F1, 1/F0±1/F1, 2/F0±1/F1, ... For the /i/ stimulus, modes occur at the periods of the crucial harmonic 2F0, which is very close to F1.

Figure 5 shows the interspike-interval representation for the high-F0 F1-continuum. As with the low-F0 stimuli, the largest modes in the pooled interval distribution for the high-F0 stimuli are always at the fundamental periods (n/F0). However, in contrast to the low-F0 case, secondary modes always occur at the periods of a higher crucial harmonics (1/2F0). Thus, individual harmonics, rather than formant frequency, are explicitly represented in the pooled interval distribution for high F0s, when harmonics are resolved in rate-place profiles.

Figure 6 shows interspike-interval representation for the high-F0 L(F0)-continuum. Pooled interspike-interval distributions for the high-F0 L(F0)-continuum are very similar to those for the high-F0 F1-continuum. In both cases, modes of the pooled distribution always occur at the periods of crucial harmonics. Thus, psychophysically-equivalent manipulations of changing the level of a single harmonic and shifting the first formant frequency produce pooled interval distributions that are also similar.

3.3 Comparison with psychophysics

A priori, we expect that the neural representation of vowel quality should covary with human vowel quality judgments, changing when perceived vowel quality changes, and remaining the same when vowel quality remains the same. Likewise, good neural correlates of phonetic boundaries should remain similar for phonetic boundaries that are observed using different stimulus manipulations. Human vowel quality judgments are expressed in terms of the percentage of /i/ judgments [2] for each stimulus in L(F0) and F1 continua. Neural measures of the amplitude ratio between crucial harmonics are expressed in terms of discharge rate ratios for the rate-place representation and in terms of interval ratios for the interval-based one. Discharge rate ratio is R(F0)/R(2F0), where R(f) is the average normalized discharge rate for CF = f. Interval ratio is I(F0)/I(2F0), where I(f) is the number of intervals at 1/f in the pooled interval distribution. Figure 7 plots these two neural measures against the human judgments that would be obtained for the same stimulus continuum. In the left panel, the discharge rate ratios for both continua are nearly the same at their respective phonetic boundaries (50% /i/ judgments). In the right panel, the same human vowel quality judgments (percent /i/) are plotted against the interval ratios for their corresponding stimuli. Again, for both continua the two curves are very close at the phonetic boundary (and elsewhere as well). Thus, the amplitude ratio of crucial harmonics provides a good acoustic correlate of the phonetic boundary, and both rate-place and interval-based neural measures

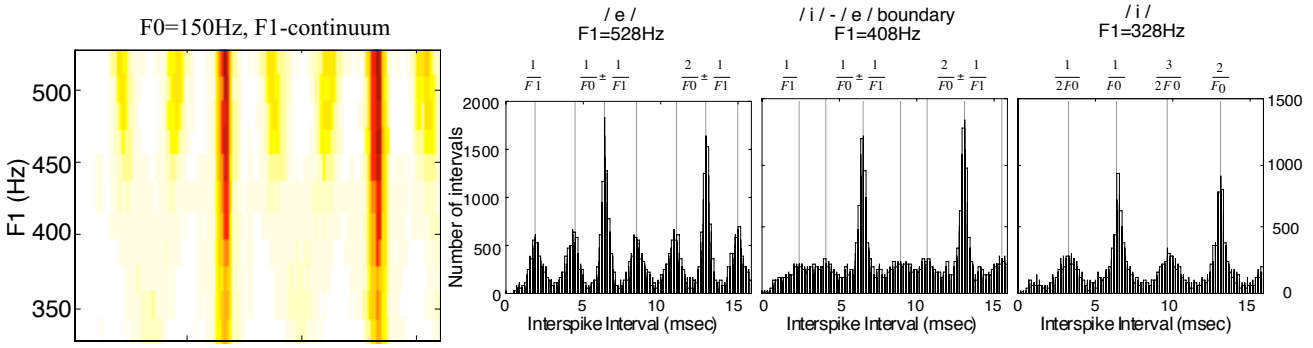

Fig. 4 The interspike-interval representation for the low-F0 (150Hz) F1-continuum. Modes of the pooled interval distribution often occur near formant period.

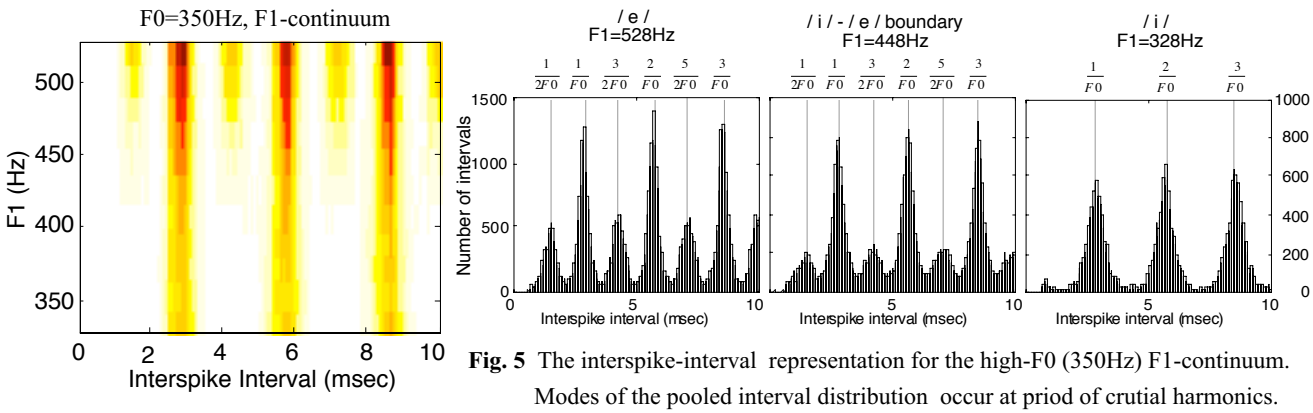

Fig. 5 The interspike-interval representation for the high-F0 (350Hz) F1-continuum. Modes of the pooled interval distribution occur at priod of crutial harmonics.

for this ratio provide correspondingly good neural correlates of the boundary.

4 SUMMARY AND CONCLUSIONS

1. Fundamental frequency plays an important role for the representation of low-frequency formants in the auditory nerve:

For high F0s, individual harmonics are physiologically resolved: peaks in rate-place profiles occur at the crucial harmonics, and modes of interval histograms are always at periods of specific harmonics. Thus, no explicit representation of formant frequency exists in the auditory nerve.

For low F0s, individual harmonics are not physiologically resolved: peaks in rate-place profiles occur near F1, and modes of interval histograms are often found near 1/F1. Thus, an explicit representation of formant frequency exists in the auditory nerve.

2. There exist correlates of the /i/-/e/ phonetic boundary in the amplitude ratios of crucial harmonics, and these amplitude ratios have clear correlates in both rate-place profiles and patterns of interspike intervals.

These results have broader implications for vowel perception by humans. Psychophysical and physiological evidence suggests that the human ear is more frequency-selective than the cat ear [6][7]. Psychophysical measures of frequency selectivity in the human are 50-100 Hz for low frequencies. In contrast, the effective bandwidths of auditory-nerve fibers in the cat exceed 150 Hz. Interpreting the cat data in the light of these species differences leads to the following conclusions:

For all voices (men, women and children), harmonics near low-frequency formants (< 1000 Hz) are likely to be physiologically resolved. Here the amplitude ratio of crucial harmonics rather than formant frequency per se may be the key cue for vowel quality. Extrapolating from the cat data, we expect that invariant correlates of the phonetic boundary exist in both rate-place and temporal discharge patterns of the human auditory nerve.

For male voices, harmonics near higher-frequency formants (>1000 Hz) are not likely to be physiologically resolved. Here, we expect that formant frequencies, rather than individual harmonics, are explicitly represented in both rate-place and temporal discharge patterns of the human auditory nerve.

ACKNOWLEDGMENT

The first author thanks Dr. Ken'ichiro Ishii for supporting him to carry out this work at EPL and NTT BRL. This research was supported by NIDCD grants DC02258 and DC00119.

REFERENCES

[1] Glasberg, B. and Moore, B. (1990): Derivation of auditory filter shapes from notched-noise data, Hearing Research 47, 103-138

[2] Hirahara, T. (1993) : On the role of relative harmonics level around the F1 in high vowel identification, ARO Abstract, 65

[3] Sachs, M.B. and Young, E.D. (1979): Encoding of steady-state vowels in the discharge patterns of auditory-nerve fibers: representation in terms of discharge rate, J.Acoust.Soc.Am., 66, 1381-1403

[4] Sachs, M.B. (1985): Speech Encoding in the Auditory Nerve, in *Hearing Science*, Edited by C.Berlin. (Taylor & Francis, London, 1985), pp.261-307

[5] Cariani, P. and Delgutte, B.:Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. Journal of Neurophysiology. (in press).

[6] Greenwood, D. (1990): A cochlear frequency-position function for several species - 29 years later, J.Acost. Soc.Am. 87, 2592-2605

[7] Pickles, J. (1979): Psychophysical frequency resolution in the cat as determined by simultaneous masking and its relation to auditory-nerve resolution, J.Acost.Soc.Am. 66, 1725-1732

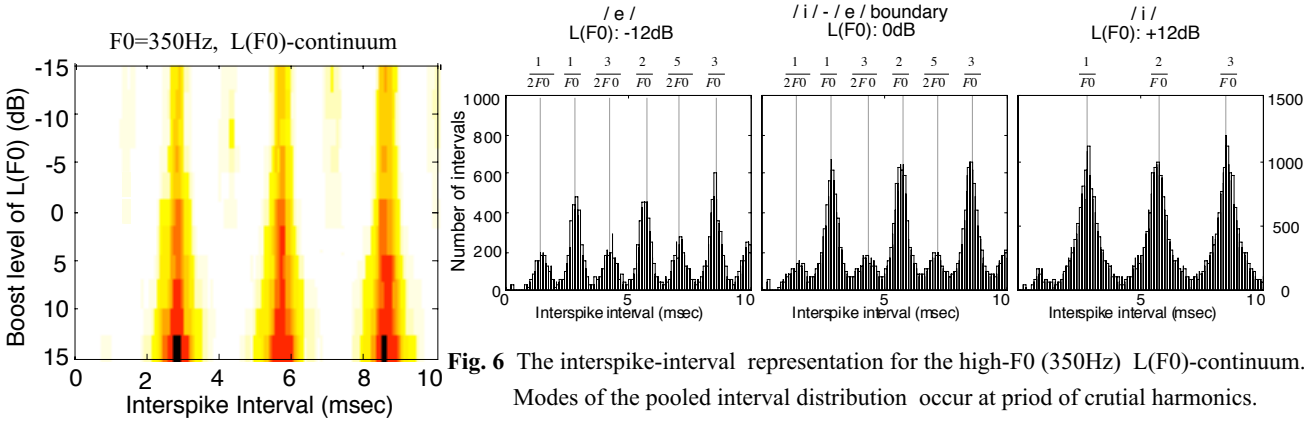

Fig. 6 The interspike-interval representation for the high-F0 (350Hz) L(F0)-continuum. Modes of the pooled interval distribution occur at priod of crutial harmonics.

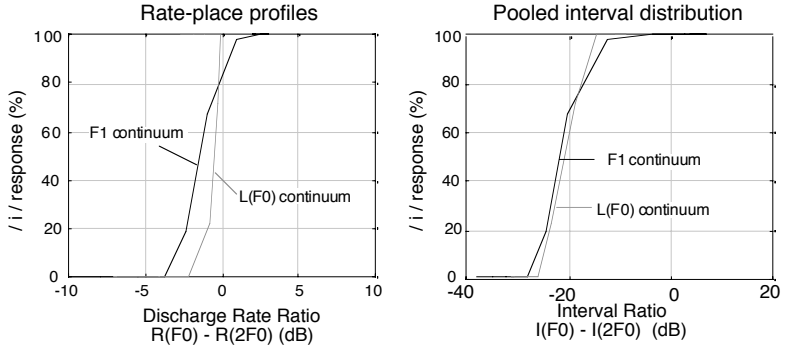

Fig. 7 The percentage of /i/ judgements for each stimuli in F1- and L(F0)-continuum is plotted against two measures of neural response, rate-place measure (left) and interspike interval measure (right). Neural correlates of the /i/-/e/ boundary are invariant with respect to stimulus manipulations that do not alter the boundary.

Temporal Coding of Periodicity Pitch in the Auditory System: An Overview

Peter Cariani

Eaton Peabody Laboratory, Massachusetts Eye and Ear Infirmary, Department of Otology and Laryngology, Harvard Medical School, Boston, Massachusetts, USA

SUMMARY

This paper outlines a taxonomy of neural pulse codes and reviews neurophysiological evidence for interspike interval-based representations for pitch and timbre in the auditory nerve and cochlear nucleus. Neural pulse codes can be divided into channel-based codes, temporal-pattern codes, and time-of-arrival codes. Timings of discharges in auditory nerve fibers reflect the time structure of acoustic waveforms, such that the interspike intervals that are produced precisely convey information concerning stimulus periodicities. Population-wide inter-spike interval distributions are constructed by summing together intervals from the observed responses of many single Type I auditory nerve fibers. Features in such distributions correspond closely with pitches that are heard by human listeners. The most common all-order interval present in the auditory nerve array almost invariably corresponds to the pitch frequency, whereas the relative fraction of pitch-related intervals amongst all others qualitatively corresponds to the strength of the pitch. Consequently, many diverse aspects of pitch perception are explained in terms of such temporal representations. Similar stimulus-driven temporal discharge patterns are observed in major neuronal populations of the cochlear nucleus. Population-interval distributions constitute an alternative time-domain strategy for representing sensory information that complements spatially organized sensory maps. Similar autocorrelation-like representations are possible in other sensory systems, in which neural discharges are time-locked to stimulus waveforms.

KEYWORDS

neural codes, interspike intervals, autocorrelation, phase-locking, temporal correlation, sensory coding, vowels, voice pitch, speech perception

THE NEURAL CODING PROBLEM

The neural coding problem, how populations of neurons represent and convey information through trains of spikes, is fundamental to our understanding how sensory systems function (Boring, 1942; Mountcastle, 1967; Perkell & Bullock, 1968; Uttal, 1973; Wasserman, 1992; Cariani, 1995; Rieke et al., 1997; Richmond & Gawne, 1998; Gerstner, 1999). The neural coding problem in perception involves mappings (Fig. 1) between stimulus, neural response, and percepts, whose biological basis can be approached from considerations of structure, function, and functional organization. These considerations involve different disciplines. Psychophysics seeks to understand the relation between stimulus and percept. Currently most neuroscience research is devoted to understanding the structure-function relationship of neurons on both the molecular and cellular levels. At the neural systems level, most current sensory neurophysiology focuses on understanding the relation between stimulus and neural response (system identification). Although

Reprint address:
243 Charles St., Boston MA 02114 USA
fax: +1- (617) 720-4408;
email: peter@epl.meei.harvard.edu

a great deal is known about neural response properties at many levels of the auditory system, we do not yet have firm understandings of which particular response properties subserve the perceptions of auditory-form qualities, such as pitch, timbre, consonance, and phonetic identity. For auditory forms, solution of the neural coding problem entails identifying which aspects of the neural response are responsible for the perceptual detections, discriminations, and recognitions that can be realized by the system as a whole. In semiotic terms, neural responses shorn of their functional roles are signs, whereas neural codes and representations constitute those aspects of the neural responses that have particular functional, informational significance. In the auditory context, a major focus of such investigations is to find strong psychoneural correspondences between patterns of activity in auditory neurons and the auditory percepts that invariably accompany them. Once such correspondences are found, then one can posit possible neural processing strategies that can make use of such information and look in the auditory pathway for specific neural mechanisms that might subserve such processing. The ultimate goal of these efforts is to understand the biological design principles, the functional organization of the auditory system as an informational system, that are essential for its perceptual and cognitive capabilities. Neural codes, the manner in which sensory information is represented by the system, lie at the heart of this functional organization.

A number of biological and behavioral constraints narrow the search for viable candidate codes. Knowing how the system is constructed, how the elements behave and how they are interconnected, places strong constraints on how

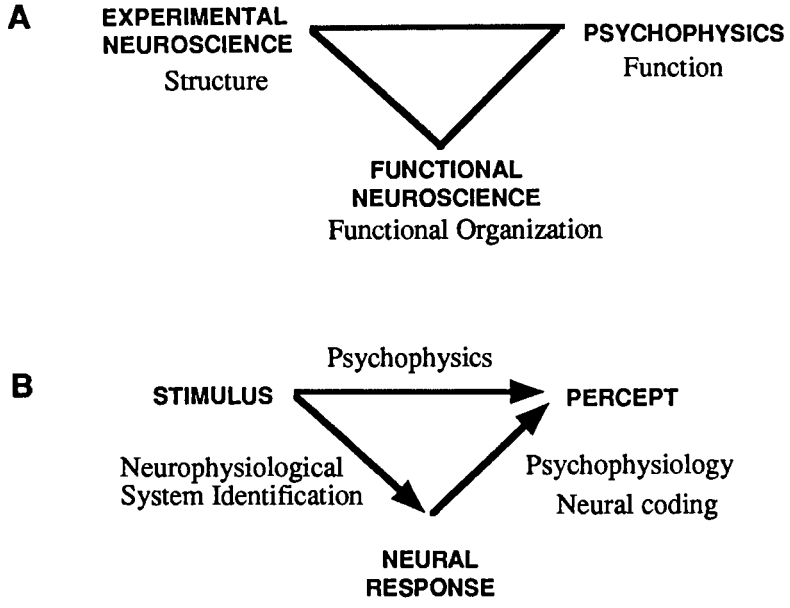

Fig. 1: Structure, function, and functional organization in the nervous system. Mappings between stimulus, neural responses, and their related percepts.

the system can handle information. Neuroanatomy supplies us with the interconnections, neurophysiology with the response properties of the parts, and molecular and cellular neuroscience with more detailed understanding of their operation. Similarly, knowing what perceptual functions the system can and cannot perform imposes a different set of functional constraints. Here information-theoretic approaches have been used to quantify how much information about the stimulus can be extracted from neural responses under particular coding schemes (Bialek et al., 1991; Rieke et al., 1997; Richmond & Gawne, 1998). Decision-theoretic approaches have been used to test how well neural information represented via particular coding schemes covaries with perceptual capabilities (Siebert, 1968; Siebert, 1970; Srulovicz & Goldstein, 1977; Delgutte, 1995). For example, decision-theoretic criteria can use the high precisions of perceptual discriminations that the system can perform under challenging conditions to narrow down the field of candidate codes. Potential codes are eliminated if not enough information exists in the neural response to support the observed precisions, or if the information is not present under all the confounding conditions under which the system is able to function. Strong perceptual and cognitive equivalence classes yield other clues as to the nature of the information being processed and of the modes by which it is utilized.

Neuroanatomical and neurophysiological considerations inform us as to how the parts of the system are interconnected and how they behave under particular conditions, but by themselves do not inform us as to which parts are essential for which functions, or how neural responses are interpreted by the rest of the system (Kiang, 1975). The psychological sciences inform us as to the functional capabilities of the system, but by themselves do not inform us the details of the neural mechanisms, what parts are needed, and how they must be organized to achieve perceptual functions. A complementary approach is therefore needed to focus on how the system is organized to perform its functions. Currently this approach comes under the rubric of functional, integrative, or computational neuroscience. In the context of informational functions, functional organization involves those aspects of neural responses that convey information and those aspects of neural structure that permit this information to inform behavior usefully.

Neuroanatomical and neurophysiological considerations inform us as to how the parts of the system are interconnected and how they behave under particular conditions, but by themselves they do not inform us as to which parts are essential for which functions – how neural responses are interpreted by the rest of the system (Kiang, 1975). The psychological sciences inform us as to the functional capabilities of the system, but by themselves they do not inform us the details of the neural mechanisms – what parts are needed and how they must be organized so as to achieve perceptual functions. A complementary approach is therefore needed to focus on how the system is organized so as to perform its functions. Currently this approach comes under the rubric of functional, integrative, or computational neuroscience. In the context of informational functions, functional organization involves those aspects of neural responses that convey information and those aspects of neural structure that permit this information to usefully inform behavior.

TAXONOMY OF NEURAL PULSE CODES

Many different kinds of neural pulse codes are possible (Fig. 2). Neural coding of sensory information can be based on discharge rates, interspike interval patterns, latency patterns, interneural discharge synchronies and correlations, spike-burst structure, or still more elaborate cross-neuron volley-patterns. Sensory coding can be based on the mass-statistics of many independent neural responses (population codes) or on the joint properties of particular combinations of responses (ensemble codes) (Hatsopoulos et al., 1998). Amidst the many ways that neural spike trains can convey sensory information are

fundamentally two basic ideas: "coding-by-channel" and "coding-by-time". Channel-based codes depend upon the activation of specific neural channels or of configurations of channels. Temporal codes, on the other hand, depend on the relative timings of neural discharges rather than on which particular neural channels respond how much. Temporal codes can be based on particular patterns of spikes within spike trains (temporal-pattern codes) or on the relative times-of-arrivals of spikes (time-of-arrival codes).

The three different modes of neural coding: coding by channel, coding by temporal pattern, and coding by time-of-arrival, are complementary and correspond respectively to different, independent, and general aspects of signals:

a) the physical channel through which the signal is transmitted,
b) its internal form (for example, its waveform or Fourier spectrum), and
c) its time of arrival.

The absolute magnitude of the signal constitutes a fourth, intensive aspect that can be used in conjunction with the other three. For encoding multiple kinds of stimulus properties in a signaling system, each signal requires two independent variables, signal-type and signal-value. One variable conveys the type or category of the information that is contained in the signal, whereas the other encodes the particular state of the signal amongst the possible alternative states. In artificial devices, the signal-type is most commonly conveyed by the particular channel through which a signal is sent (consider the many types of information conveyed by the respective wires leading to different gauges on the dashboard of a car). The identity of the channel conveys to the rest of the device what kind of information is being sent (namely, to which type of sensor the wire is connected). Similarly, in artifacts, signal-value is usually conveyed by the amplitude of the signal, often a voltage. Neural coding schemes similarly

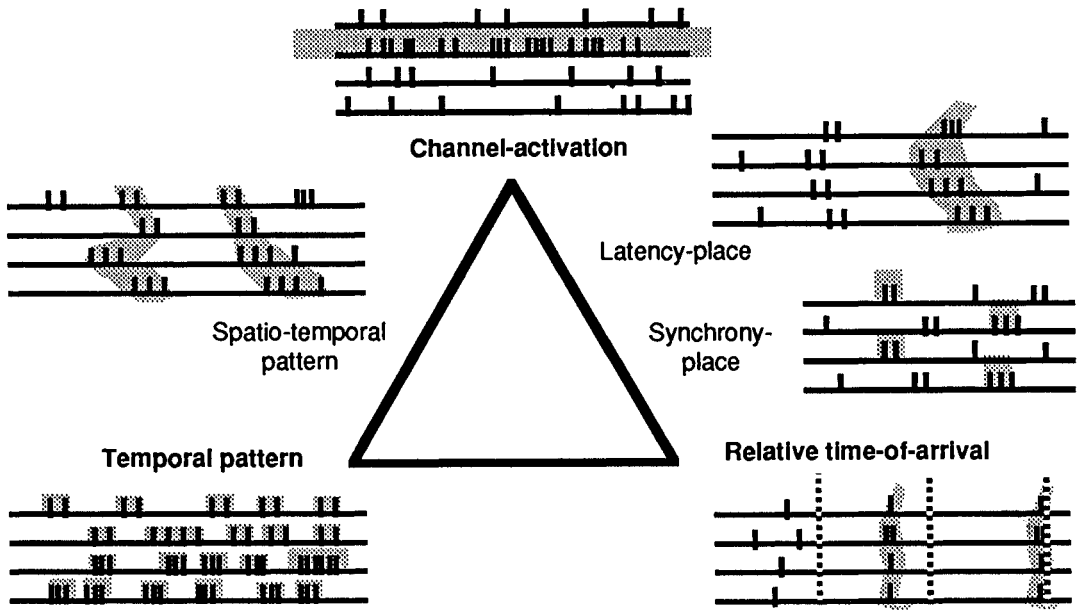

Fig. 2: A space of possible pulse codes. Three complementary modes of coding are shown at the vertices, with their combinations on the edges.

require two sets of independent coding variables. Most typically, channel-based neural-coding schemes use the identities of neurons to encode signal-types, whereas some intensive measure of activation, such as discharge rate, encodes signal-value. So constructed, channel-based coding schemes depend critically upon which particular neurons are activated how much. If the connectivities of neurons are suddenly rearranged in such a system, the coherence of neural representations will be disrupted, at least until the system can be adaptively rearranged to reflect the new channel-identities.

Many different channel-based coding schemes are possible. Such schemes can range from simple, unidimensional representations to low-dimensional sensory maps to higher dimensional feature detectors. In simple "doorbell" or "labeled line" systems, activation (or suppression) of a given neuron signals the presence or absence of one particular property. In more multipurpose schemes, neurons are differentially tuned to particular stimulus properties, such as frequency, periodicity, intensity, duration, or external location. Profiles of average discharge rates across a population of such tuned elements then convey multidimensional information about a stimulus. When spatially organized in a systematic manner by their tunings, these elements form sensory maps, in which spatial patterns of channel-activation can then represent arbitrary combinations of those stimulus properties. In lieu of coherent spatial order, tuned units can potentially convey their respective channel-identities through the specificity of spatially distributed neural connections. More complex constellations of properties can be represented via more complex concatenations of tunings to form highly specific "feature-detectors". In the absence of coherent tunings, combinations of idiosyncratic response properties can potentially form "across-neuron pattern codes" of the sort that are commonly proposed for the olfactory system. Nevertheless, idiosyncratic across-neuron patterns and associative learning mechanisms have fundamental difficulties in explaining common strong perceptual equivalence classes that are shared by most humans and are largely independent of an individual's particular history (Gesteland et al., 1965). Although these various functional organizations, from labeled lines to feature detectors to across-neuron patterns, encompass widely diverse modes of neural representation, all draw on the same basic strategy of coding-by-channel. In channel-coding schemes, it is usually further assumed that distinctions between alternative signal-states are encoded by different average discharge rates (Shadlen & Newsome, 1994; Shadlen & Newsome, 1998). The combination of channel- and rate-based coding has remained by far the dominant neural coding assumption throughout the history of neurophysiology (Boring, 1942; Barlow, 1995), and, consequently, forms the basis for nearly all our existing neural-network models.

Within channel-coding schemes, aspects of the neural response other than rate, such as relative latency or temporal pattern, can also play the role of encoding alternative signal states (for example, the latency-place and spatiotemporal codes shown in Fig. 2). In a simple latency-channel code, channels producing spikes at shorter latencies relative to the onset of a stimulus indicate stronger activation of tuned elements, which can be used to encode stimulus intensity (Stevens, 1971), location (Brugge et al., 1996), or other qualities. Common-response latency, in the form of interchannel synchrony, has been proposed as a strategy for grouping channels to form discrete, separate objects (Singer, 1990; Singer, 1995). In this scheme, rate patterns across simultaneously activated channels encode object-qualities, whereas interchannel synchronies (joint properties of response latencies) create perceptual organization, which channels combine to encode which objects. The concurrent use of multiple coding vehicles, channel, rate, and common time-of-arrival permits time-division multiplexing of multiple objects. Still, other kinds of asynchronous multiplexing are possible if other coding variables, such as common temporal pattern and phase coherence, are used (Cariani, 1997).

Characteristic temporal discharge patterns can also convey information about stimulus qualities.

Neural codes that rely predominantly on the timings of neural discharges have been found in a variety of sensory systems (Mountcastle, 1967; Perkell & Bullock, 1968; Chung et al., 1970; Kozak & Reitboeck, 1974; Covey, 1980; Emmers, 1981; Bialek et al., 1991; Carr, 1993; Cariani, 1995; Cariani, 1997). Conceptually, these temporal codes can be divided into time-of-arrival and temporal-pattern codes.

Time-of-arrival codes use the relative times of arrival of spikes in different channels to convey information about the stimulus. Examples of time-of-arrival codes are found in many sensory systems that utilize the differential times of arrival of stimuli at different receptor surfaces to infer the location of external objects (Bower, 1974; Carr, 1993). Strong examples are auditory localizations that rely on the time-of-arrival differences of acoustic signals at the two ears, echolocation range-findings that rely on time-of-arrival differences between emitted calls and their echoes, and electroceptive localizations that use the phase-differences of internally generated weak electric fields at different locations of the body to infer the presence of external phase distortions caused by nearby objects.

Temporal pattern codes, such as interspike interval codes, use temporal patterns between spikes to convey sensory information. In a temporal pattern code, the internal patterns of spike arrivals bear stimulus-related information. The simplest temporal pattern codes are interspike interval codes, in which stimulus periodicities are represented using the times between spike arrivals. More complex temporal pattern codes use higher-order time patterns consisting of interval sequences (Emmers, 1981; Lestienne & Strehler, 1987). Like time-of-arrival codes, interval and interval-sequence codes could be called correlational codes because they rely on temporal correlations between individual spike-arrival events. These codes should be contrasted with conceptions of temporal coding that rely on temporal variations in average discharge rate or discharge probability (for example, Richmond & Gawne, 1998), which count numbers of events across stimulus presentations as a function of time and then perform a coarse temporal analysis on event-rates. Both time-of-arrival and temporal-pattern codes depend on the stimulus impressing itself, in one way or another, on the timings of neural discharges. The stimulus-related temporal discharge patterns, on which temporal-pattern codes depend, can arise in two ways, through stimulus-locking and through intrinsic-time courses of response.

For stimulus-locking, the discharges of sensory neurons follow the time-amplitude course of the stimulus waveform. The highly stimulus-locked nature of discharges in the auditory nerve and the cochlear nucleus is evident in the peri-stimulus time histograms shown in the figures below. As long as a monotonic relation exists between the amplitude of the driving input and the probability of subsequent discharge, temporal correlations will be produced between waveform and spike train. In the auditory system, as in many other sensory systems, receptor cells depolarize when stereocilia are deflected in a particular direction, such that the timings of spikes predominantly occur during one phase of the stimulus waveform as it presents itself to the individual receptor (for example, after having been mechanically filtered by the cochlea). This form of stimulus-locking is known as *phase-locking*. In the auditory system, depending upon the species, strong phase-locking can exist up to several kHz, dramatically declining as progressively higher frequencies are reached. Such phase-locking is also found in many other sensory systems, albeit usually at much lower frequencies. Phase-locked responses exist

a) to flutter-vibrations of the skin in mechanoception (Mountcastle, 1993),
b) to accelerations in the vestibular system,
c) to drifting gratings and flickering lights in the visual system (Pollen et al., 1989),
d) to inhalation cycles and odor pulses in olfaction (Macrides & Chorover, 1972; Onoda & Mori, 1980; Marion-Poll & Tobin, 1992),
e) to self-produced electrical oscillations in electroception, and
f) to the movements of muscles in proprioceptive stretch receptors.

A generalization can be made that every sensory system will show phase-locked responses to its adequate stimulus, provided that the contrast is sufficient to create distinguishable, phase-dependent responses and that modulations are slow enough for phase-dependent responses to be separated temporally. To the extent that phase-locking exists, then the time intervals between the spikes that are produced (interspike intervals) reflect stimulus periodicities, such that time intervals themselves can serve as neural representations of stimulus form. In addition, phase-locked discharges register the arrival times of nonperiodic, transient, and unitary events, such that comparisons of the arrival times of the same event at different sensory surfaces (for example, the differential time-of-arrival of an acoustic wavefront at the two ears) can serve as neural representations of stimulus location relative to those sensory surfaces.

The intrinsic temporal patterns of neural response can also convey information concerning stimulus qualities. Such temporal response patterns can be characteristic of particular receptor types, individual neurons, local neural circuits, or even whole neural populations. Stimulus-related temporal discharge patterns that are not directly locked to the time structure of the stimulus have been observed in many sensory systems: olfaction (Kauer, 1974; Macrides, 1977; Marion-Poll & Tobin, 1992; Laurent & Naraghi, 1994); gustation (Covey, 1980; Di Lorenzo & Hecht, 1993); spatial vision (Richmond et al., 1987; Richmond & Gawne, 1998); color (Kozak & Reitboeck, 1974; Wasserman, 1992). In some sensory modalities, temporal patterns of electrical stimulation appear to produce particular sensory qualities, such as taste and color (Young, 1977; Covey, 1980; Di Lorenzo & Hecht, 1993), suggesting that the temporal patterns themselves may be the neural-coding vehicles that subserve these particular qualities. Stimulus-triggered intrinsic temporal patterns that are associated with conditioning and perceptual expectations have also been found in cortical regions (John & Schwartz, 1978; John, 1990).

How might such intrinsic time patterns represent combinations of stimulus properties? One possibility is that the relative occurrence of different time patterns, associated with characteristic impulse- or step-responses of particular neurons, can serve as markers that indicate the activation of particular sub-populations of neurons. Mixtures of odorants, tastants, and wavelengths of light would then produce mixtures of the respective temporal spike patterns of the receptors and neurons that they preferentially activate. As in the population-interval representation for pitch discussed below, patterns that are associated with the individual constituents, their interactions, and their fusions presumably would exist in the population time structure. These features could then be used to discriminate basic stimulus properties and to represent mixtures.

For different stimulus-receptor combinations, many ionic and molecular mechanisms in sensory receptors are available to produce differential kinetics of activation, inactivation, and recovery. In neural populations, temporal dynamics of excitation and inhibition could similarly produce characteristic temporal patterns. Both stimulus-locked and stimulus-triggered intrinsic temporal response patterns can be found throughout the auditory pathway. Extrinsic stimulus-locked patterns are most apparent in the lower stations, whereas intrinsic patterns become more apparent as one progresses to more central stations, where neural responses become increasingly dominated by the recent history of the system as a whole.

Finally, yet another dimension to neural codes involves the joint response properties of multiple neurons. This dimension is the distinction between population codes and ensemble codes (Deadyler & Hampson, 1995; Hatsopoulos et al., 1998), between statistical orders and switch-boards (John, 1972). To represent information, population codes use the mass statistics of stimulus-driven response properties of individual, largely independent units. Examples of such population codes are population-rate vectors in the motor cortex (Lee et al., 1998), or the auditory population-interval distributions presented below. In population codes, interdependencies between

the responses of particular neurons are themselves irrelevant to the representation. Ensemble codes, on the other hand, use these interdependencies rather than common, stimulus-driven statistical structure to represent information. Response interdependencies can be reliably produced by specific interneural connectivities and time-delays. The resulting stimulus-related intrinsic correlations between the neuronal channels that are activated and/or synchronized, as well as between the latencies of spikes produced by different neurons, then can convey information about a stimulus. Perceptual grouping by means of channel-synchronizations that are not stimulus-driven would be an example of an ensemble code, in which statistics of channel activations by themselves would be insufficient for its interpretation; one would have to know which combinations of channels are synchronized at each moment. In the context of sensory coding, the relative merits of the stimulus-driven, mass-statistics of population-codes versus the stimulus-triggered, joint-response properties of ensemble codes remain to be more fully explored.

THE NEURAL BASIS OF PITCH PERCEPTION

The nature of the neural codes that subserve auditory perception have generated lively ongoing discussion and debate for most of the last 150 years (Boring, 1942; Wever, 1949; de Boer, 1976). For the most part, this discussion has been focused on whether frequency is represented (a) via rate-place codes, namely, neural discharge-rate profiles in auditory frequency maps, or (b) via temporal codes, namely, interspike interval distributions (Siebert, 1970; Moore, 1997; Evans, 1978s). In many debates about neural coding, pitch has played a pivotal role mainly because pitch is a perceptual correlate of frequency (Boring, 1942; de Boer, 1976). At the same time, pitch is also a perceptual correlate of periodic waveforms, whether single pure tones or complexes consisting of many harmonically related partials. Operationally, pitch is defined as the frequency of a pure tone to which a given sound can be reliably matched. The percept provides a very rich test bed for understanding many aspects of perception. Like color, pitch is metameric; the same pitch can be evoked by many different stimuli that can differ markedly in their power spectra. When harmonically related partials are sounded together, strong pitches at their common, fundamental frequency (F0) can be produced in the absence of any spectral component at that frequency (specifically, the "missing fundamental" is heard). These pitches are often called "low" pitches because the fundamental has a lower pitch than those that are associated with any of the individual partials. Such pitches are often called "periodicity" pitches because the low pitch at the fundamental reflects the periodicity of the recurring time pattern that is associated with the whole harmonic complex. Thus, combinations of partials give rise to new low pitches that are not heard in the separate constituents. Pitches produced by such complex tones are consequently "emergent" perceptual Gestalts, products of the relations between parts rather than of the parts themselves. Finally, pitch is largely invariant with respect to a host of factors, such as stimulus intensity and location, that produce large changes in the responses of auditory neurons. These perceptual invariances focus the search for the neural basis of pitch on the aspects of neural activity displaying similar stability.

Historically, a strong case for temporal coding of pitch has always been made (Troland, 1929; Boring, 1942; Wever, 1949), although the pendulum of scientific opinion has swung back and forth between spectral pattern and temporal theories several times now (de Boer, 1976; Evans, 1978; Lyon & Shamma, 1995). Although auto-correlation-based models for pitch were first proposed almost a half-century ago (Licklider, 1951; Licklider, 1959), only during the last two decades have similar kinds of global, interval-based models been revived and extended (van Noorden, 1982; de Cheveigné, 1986; Meddis & Hewitt, 1991; Slaney & Lyon, 1993; Lyon & Shamma, 1995; Cariani & Delgutte, 1996a, 1996b; Meddis & O'Mard, 1997; Moore, 1997).

In physiological studies at the level of the auditory nerve of the cat (Cariani & Delgutte, 1996a, 1996b), a robust and pervasive correspondence was found between interspike interval statistics of populations of auditory nerve fibers and pitches that are produced by a wide array of complex tones. The auditory nerve is a strategic location for the study of pitch, the conduit through which must pass virtually all the auditory information that the central auditory system uses for the representation of sounds. Thus, whatever the nature of subsequent processing, the necessary information for all auditory capabilities must be present in the responses of auditory nerve fibers. For this reason, the auditory nerve has been one of the most intensively studied neural populations in the nervous system (Kiang et al., 1965; Rose et al., 1967).

METHODS

The auditory nerve responses presented here come from the same data set that has been published previously in (Cariani & Delgutte, 1996a; 1996b), where experimental methods, stimuli, and analytical procedures are described in detail. Briefly, stimuli were numerically synthesized and delivered via closed, calibrated acoustic systems to Dial-anesthetized cats with normal hearing. Posterior craniectomy and partial retraction of the cerebellum permitted the visually-guided insertion of glass microelectrodes into the auditory nerve near the internal auditory meatus. The auditory nerve in the cat consists of two populations of spiral ganglion afferents: myelinated Type I radial afferents (90% to 95%) and unmyelinated Type II outer spiral afferents (5% to 10%) (Ryugo, 1992). The responses of single Type I auditory nerve fibers were recorded serially, using standard electrophysiological techniques. For each fiber, the characteristic frequency (CF), the discharge-rate response threshold, and the spontaneous rate were measured. The CF is the frequency to which a fiber has its lowest sound-pressure threshold (namely, the frequency for which the lowest sound-pressure level reliably elicits an extra spike within a 50 ms period). Characteristic frequencies therefore provide an indication of the cochlear "place" from which an auditory nerve fiber receives its synaptic inputs. Units in the cochlear nucleus were recorded extracellularly, using tungsten electrodes positioned under direct visual guidance.

NEURAL CORRELATES OF PITCH IN THE AUDITORY NERVE

In these studies, microelectrode recordings were made of responses of single auditory nerve fibers to stimuli that produce low, periodicity pitches in humans. Figure 3 shows the responses of 51 auditory nerve fibers to 100 presentations of such a stimulus.

The waveform, power spectrum, and autocorrelation function of the vowel stimulus are shown in panels 3A,C,E. The vowel is a harmonic complex whose partials are all integer multiples of its fundamental frequency (F0=80 Hz) and whose waveform is periodic, repeating every fundamental period (1/F0=12.5) ms. Perceptually, the vowel produces a strong low pitch at its fundamental frequency (F0=80 Hz), whereas the vowel quality or timbre is determined by its single, formant frequency (F1=640 Hz) and its bandwidth (50 Hz). The temporal patterns that are associated with the fundamental and the formant can be seen in the waveform (3A) of the vowel, whereas their respective harmonic spacings and concentration of energy in the formant can be seen in the power spectrum of the vowel (3C). The vowel stimulus was delivered at a moderate level (60 dB SPL).

Response peristimulus time histograms (PSTHs) for the whole ensemble of fibers are shown in Fig. 3B. The PSTHs are ordered by their respective characteristic frequencies (CFs). Immediately striking is the wide extent to which stimulus-driven temporal discharge patterns predominate over the entire auditory nerve array. Periodicities related to the fundamental F0, and hence, to the pitch period, are distributed across the entire array in the responses of fibers, with CFs ranging from 200 Hz to over 10 kHz. Given

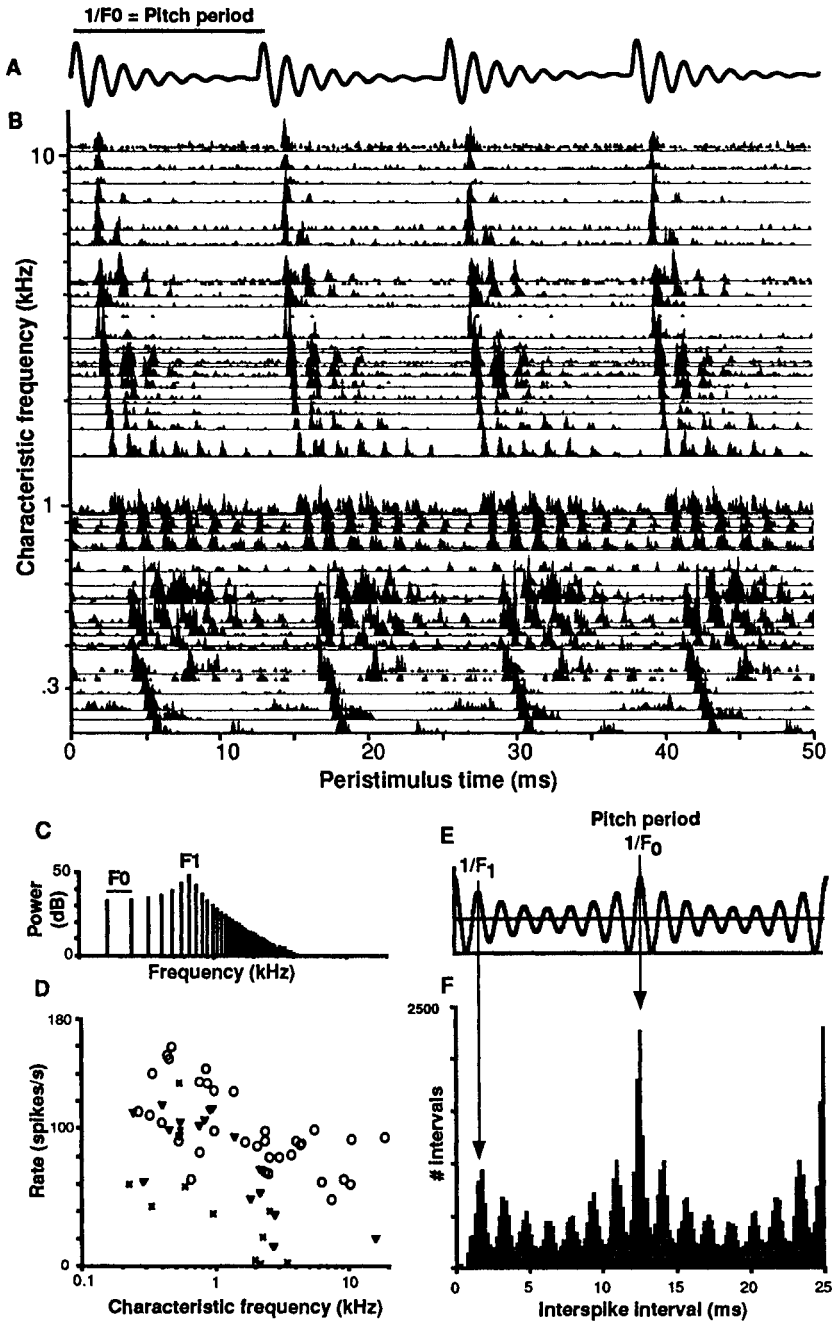

Fig. 3: Auditory nerve response to a single-formant vowel. A. Vowel waveform. A strong, low voice pitch is heard corresponding to the fundamental period, 1/F0=12.5 ms. B. Peristimulus time histograms of 51 cat auditory nerve fibers to 100 presentations at 60 dB SPL, arranged by characteristic frequency (CF). C. Power spectrum of the vowel (logarithmic frequency scale). The fundamental frequency, (F0=80 Hz = frequency spacing of harmonics) and the formant frequency (F1 = 640 Hz = spectral peak) are indicated. Bandwidth = 50 Hz. D. Discharge rates as a function of CF and spontaneous rate (SR) (circles, high SR; triangles, medium SR; crosses, low SR). E. Stimulus autocorrelation function. Arrows indicate formant period 1/F1 and fundamental period 1/F0. F. Population-interval distribution formed by summing all-order intervals from all fibers.

that the stimulus has relatively little power above 1 kHz (Fig. 3C), this result is perhaps even more remarkable. To a greater or a lesser degree, all temporal discharge patterns follow the stimulus waveform, reflecting the relation between the respective fiber CFs and the stimulus spectrum. The reason for this near ubiquity of common temporal structure lies in the broadband nature of the responses at moderate- and high-stimulus levels and in the frequency asymmetry of cochlear tuning. The broad, low-frequency tails of tuning curves are such that low-frequency components presented at moderate levels (>50 dB SPL) can weakly drive large numbers of auditory nerve fibers whose CFs are well above them (Kiang & Moxon, 1974; Kim & Molnar, 1979). Discharge rates as a function of CF and spontaneous rate are shown in Fig. 3D for a rough comparison with the stimulus power spectrum. Spectral pattern representations for pitch that are based on discharge rates would require that

a) the individual harmonics be separated in population-rate profiles,
b) their frequencies be associated with the individual harmonics estimated, and
c) their harmonic relations be analyzed to infer the frequency of their common fundamental.

The two dominant periodicities of the vowel, F0 and F1, can be readily seen in the discharge patterns of fibers in different CF regions. At this sound-pressure level, intervals related to the fundamental are found virtually everywhere, whereas formant-related periodicities are concentrated primarily in the CF regions that are nearest to the formant. More detailed views of the responses of two fibers with different CFs are shown in Fig. 4. A fiber whose CF is in the formant region (CF=950 Hz, Fig. 4A to 4F) discharges throughout most of each vowel period. A second fiber whose CF is above the formant region (CF=2100 Hz, Fig. 4H to 4K) responds less vigorously to the stimulus, producing spikes mostly at the onset of the vowel period. In Fig. 4. first-order and all-order interspike interval histograms are shown for the two fibers. A first-order interval histogram (Fig. 4E and 4J) tabulates the distribution of interspike intervals between consecutive spikes, whereas an all-order interval histogram (4F and 4K), also called an autocorrelation histogram, tabulates the distribution of intervals between both consecutive and nonconsecutive spikes. Both fibers produce intervals that are related to the fundamental period (1/F0=12.5 ms) and to components in the formant region (1/F1=1.6 ms), albeit in different proportions.

It should be noted here that some measures that have been traditionally used to quantify temporal structure in neural responses, such as first-order interval distributions, period histograms, and synchronization indices, can provide misleading comparisons. For example, the discharge rate of the higher-CF fiber is more highly modulated, so that its period histogram would show spikes that are distributed over a smaller fraction of the vowel period, producing a correspondingly higher synchronization index. The higher-CF fiber might therefore be thought to better encode the fundamental period. Similarly, the higher-CF fiber produces more first-order intervals at the fundamental period than does its formant-region counterpart. But nevertheless, in absolute terms, the formant-region fiber contributes more all-order, F0-related interspike intervals to the population response. The reason for the inversions concerns the relative nature of these measures; for both measures, adding extra, intervening spikes alters the apparent amount of F0-related temporal structure. Because synchronization indices are relative, vectorial additions, adding extra spikes throughout the period, degrades the index. Because first-order interval distributions omit longer intervals when intervening spikes are present, these distributions systematically exclude longer F0-related intervals as discharge rates increase. As discharge rates generally increase with the level, if first-order intervals were used, then the neural representation of low fundamental frequencies should have worsened at higher levels, a trend that is not observed in the psychophysics. By contrast, all-order intervals that are associated with particular periodicities are not adversely affected by the extra, intervening spikes; hence such intervals

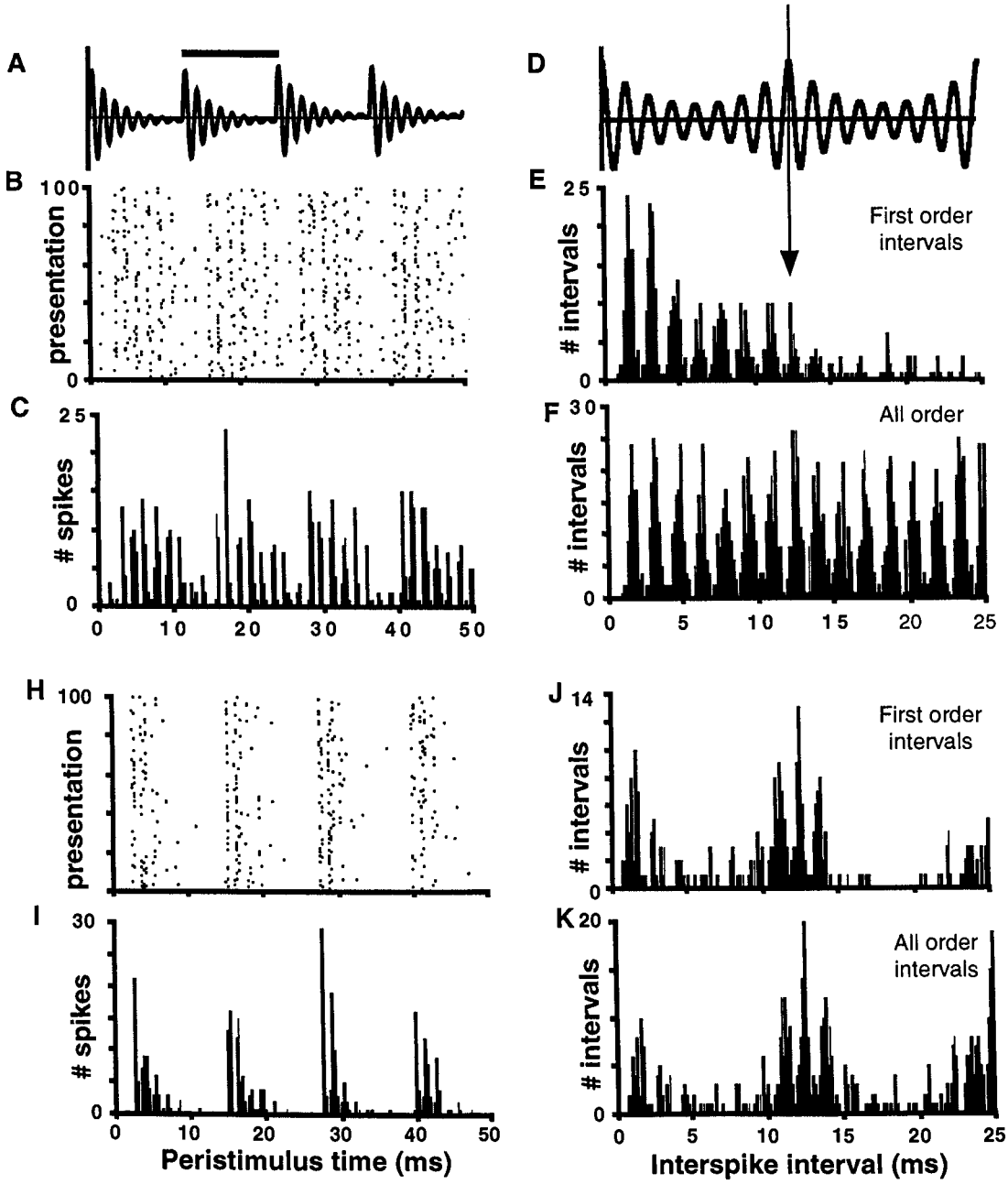

Fig. 4: Responses of two auditory nerve fibers with different CF's to 100 presentations of a single formant vowel, 60 dB SPL. A-F. Unit 25-19, CF = 950 Hz, near the formant region. H-K. Unit 25-91, CF = 2.1 kHz, well above the formant region. A. Vowel waveform. Fundamental period 1/F0 (line) is 12.5 ms, F0 = 80 Hz. B. Dot raster display of individual spike arrival times. C. Peristimulus time histogram of spike arrival times. D. Stimulus autocorrelation function. Vertical line indicates fundamental period, 1/F0 = voice pitch. E. Histogram of first-order interspike intervals (between consecutive spikes). Arrows indicate intervals near the fundamental/pitch period. F. Histogram of all-order intervals (between both consecutive and nonconsecutive spikes). H-K. Corresponding histograms for the second fiber.

constitute a neural coding strategy that better mimics perception in its behavior. For these reasons, it is important to choose measures of temporal response structure appropriate to the kinds of neural codes that one is investigating. Every neural response measure that one analyzes carries with it an implicit neural coding hypothesis.

Population-interval distributions are formed by summing together the all-order interspike interval distributions of individual fibers (Fig. 3F). Population-interval distribution serves as a rough estimate of the interval statistics of the entire auditory nerve. Because this distribution is the sum of many autocorrelation histograms or channel-auto-correlations such a distribution is often called the "summary autocorrelation" in many signal-processing and auditory simulations contexts (Meddis & Hewitt, 1991; Lyon & Shamma, 1995). The most salient aspect of this distribution is the large major peak associated with the fundamental period (1/F0=12.5 ms) which, in turn, corresponds to the low pitch that is heard. For harmonic stimuli, all-order intervals at the fundamental period are always at least as numerous as those associated with any other periodicity (Rose, 1980), so that invariably, when the all-order intervals from many fibers are pooled together, the intervals at the fundamental are the most abundant. The second major peak, at 25 ms, is also associated with the fundamental: the second peak corresponds to two fundamental periods. These major interval peaks correspond to the major peaks in the stimulus autocorrelation function (Fig. 3E). Thus, the most common interspike intervals that are generated at the level of the auditory nerve correspond to the pitch of the stimulus. This concordance was found to be the case for a wide range of fundamental frequencies and for many other kinds of harmonic stimuli as well, using both neurophysiological data (Cariani & Delgutte, 1996a; 1996b) and auditory nerve simulations (Meddis & Hewitt, 1991; Meddis & O'Mard, 1997).

Yet another salient aspect of the population-interval distribution (Fig. 3F) is its similarity in form with the autocorrelation function of the stimulus (Fig. 3E). The similar locations of major and minor peaks in both distributions is a general consequence of phase-locking, namely, temporal correlation between the stimulus waveform and the spike timings. In effect, population-interval distributions can serve as autocorrelation-like representations of the stimulus that contains the same information, up to the frequency limits of phase-locking, as its power spectrum. Thus, operations that are formally related to Fourier analysis can be neurally realized in the time domain by using all-order interval distributions.

Interspike interval information is extremely precise, permitting the fundamental period to be reliably estimated with a high degree of accuracy. From the responses of a few thousand spikes, estimates of the fundamental frequencies of stimuli producing strong pitches, such as the single-formant vowel, typically have standard errors on the order of 1%. This estimate can be compared with the ability of human listeners (~30,000 Type I auditory nerve fibers) to distinguish fundamental frequencies differing by fractions of a percent (cf. Siebert, 1968; Siebert, 1970).

Many other aspects of pitch perception can be explained in terms of population-interval representations. Some of these are summarized in Fig. 5, with their associated population-interval histograms. Harmonic complexes lacking frequency components at their fundamentals, such as the AM tone in Fig. 5A, nevertheless evoke strong pitches at their "missing" fundamentals. The power spectrum of the AM tone in the second plot shows the frequencies of its three components (solid lines at 480, 640, and 800 Hz) and the frequency of the low pitch heard at the fundamental (dotted line at 160 Hz). Both the stimulus autocorrelations and the population-interval distributions produced by such stimuli (rightmost plots) exhibit major peaks that correspond to these emergent pitches.

Different kinds of stimuli can give rise to the same low pitch. In one way or another, the auditory system creates strong perceptual equivalence classes for pitch. Population-interval distributions for four stimuli are shown in Fig. 5B. Despite very

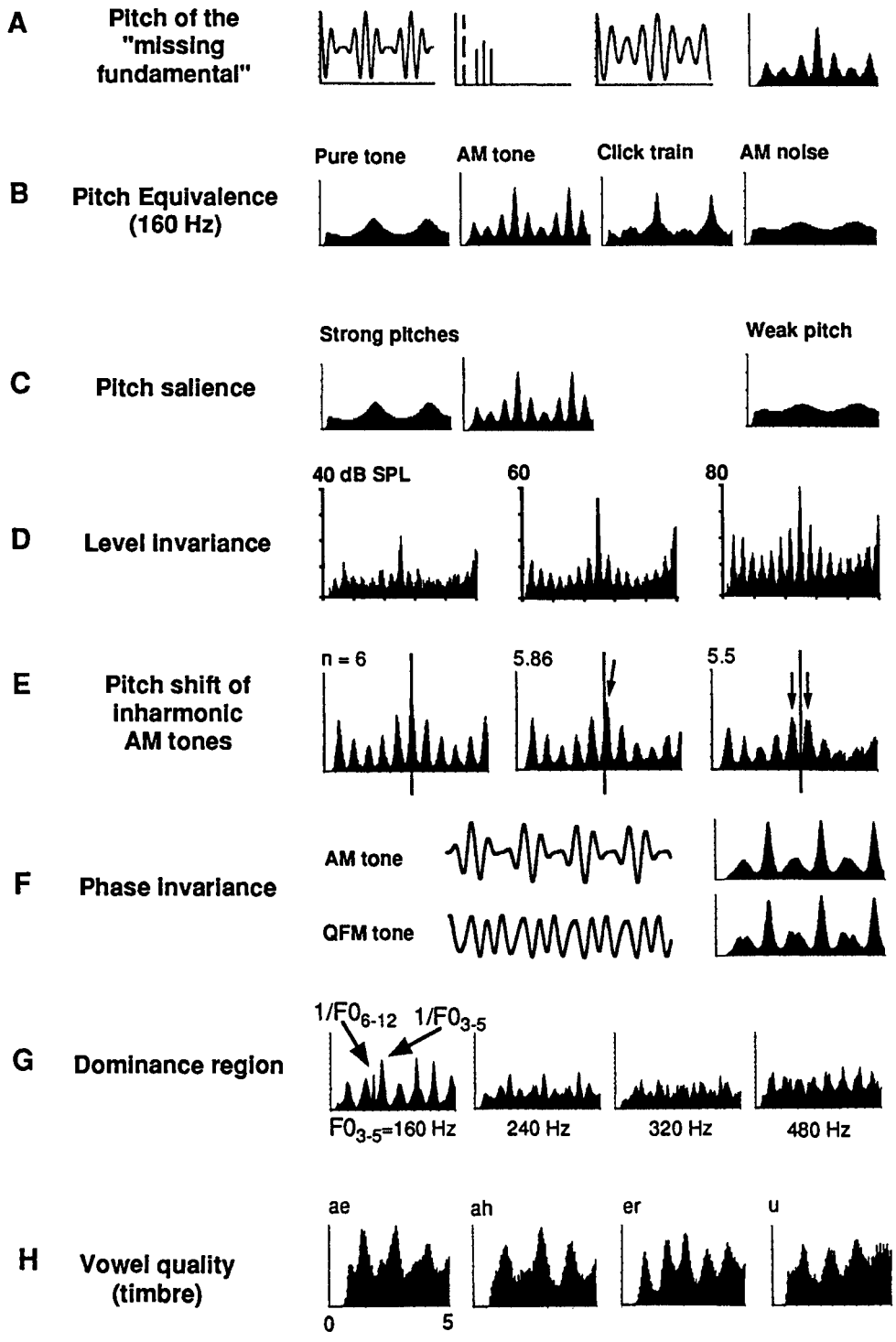

Fig. 5: Schematic summary of major correspondences between pitch percepts and population interval distributions at the level of the auditory nerve. The population-interval histograms plot relative numbers of all-order intervals (ordinates) of different durations (abcissas). Interval ranges for the histograms: 0-5 ms (H), 0-10 ms (A, F); 0-15 ms (B, C, E, G); 0-25 ms (D). Waveform segments are 20 ms long. See text for discussion.

different power spectra, each kind of stimulus evokes a common low pitch at 160 Hz. In all cases, the positions of the major interval peaks correspond to the common pitch period (6.25 ms). Thus, if the auditory system carried out an analysis of population-interval distributions, with the predominant interval corresponding to the pitch, then the pitch-equivalence of these stimuli would be a direct consequence of the basic neural-coding mechanisms that are used by the auditory system.

Different stimuli also differ in pitch salience, evoking stronger or weaker pitches. The population-interval distributions for three stimuli differing in pitch salience are shown in Fig. 5C. The two leftmost stimuli, a pure tone and an AM tone, evoked strong pitches, whereas the rightmost stimulus, an amplitude-modulated broadband noise, evoked a weak pitch. Qualitatively, the stimuli evoking strong pitches produced population-interval distributions with higher peak-to-mean ratios, namely, a higher fraction of all the pitch-related intervals that were produced. The stimuli producing weak pitches had low peak-to-mean ratios that were much closer to unity. The correspondence between pitch salience and peak-to-mean ratios is rough only because the pure tone produces a pitch that is always at least as salient as an equivalent AM tone, yet the peak-to-mean ratio of the AM tone was substantially greater than that of the pure tone.

The low pitches of complex tones are highly invariant with respect to stimulus intensity. Population-interval distributions for the single-formant vowel discussed above are shown in Fig. 5D for three sound pressure levels: low (40 dB SPL), moderate (60 dB SPL) and high (80 dB SPL). Like human pitch judgments, the pitches that were estimated from population-interval distributions changed very little over the 40 dB range. Similarly, the representation of formant-related periodicities remained very stable over that range. In the auditory system, such stability makes for extremely robust representations of pitch and timbre that do not degrade at moderate and high levels. In contrast, the saturation of discharge rates at these levels (Kim & Molnar, 1979), with the consequent loss of representational contrast and precision, poses fundamental problems for rate-place coding of these qualities.

Population-based correlational representations of loudness are conceivable. As stimulus levels increase, population interval distributions more closely resemble the stimulus autocorrelation: the correlation coefficient r between the stimulus autocorrelation function of the single formant vowel, and its respective population-interval histogram is 0.62 (n=17 fibers) for 40 dB SPL; rising to 0.70 (n=61 fibers) at 60 dB SPL, and 0.77 (n=31 fibers) at 80 dB SPL. The correlation coefficient is, in effect, a measure of the amount of the common stimulus-driven time structure in the neural population. These comparisons are tentative because little overlap exists among the three sets of fibers. Nevertheless, such comparisons suggest a straightforward interpretation. As stimulus levels increase, a progressively greater fraction of discharge timing is stimulus-driven, such that the ratio of stimulus-driven intervals versus uncorrelated, spontaneously produced intervals steadily increases. Thus, the loudness of an auditory object potentially could be encoded by the fraction of the common temporally structured activity with which it is associated. Such a correlational representation would effectively use the entire dynamic range of the whole auditory nerve array. In such a scheme, spontaneous activity increases the dynamic range of the system by providing an uncorrelated noise source that can be successively displaced by stimulus-driven interspike intervals.

Complex pitch phenomena can also be explained in terms of population-interval distributions. Whereas periodic, harmonic tone complexes generally evoke unambiguous low pitches, inharmonic complexes can evoke ambiguous, multiple low pitches and small pitch shifts. A half-century ago, Schouten and deBoer (deBoer, 1976) conducted a classic set of experiments to determine whether pitch perception relies on spacings between adjacent frequency components (or equivalently on waveform envelope periods) rather than on harmonic relationships between components (or equivalently on the waveform fine structure). An AM tone consists of a complex comprising three successive harmonics that evokes a clear,

unambiguous pitch at its (missing) fundamental frequency. When all three harmonics were shifted either upward or downward in frequency by the same amount, while keeping their frequency-spacings constant, the low pitch of the complex first shifted slightly by a much smaller amount than this frequency difference, an amount that was related to harmonic structure. When the frequencies were further shifted, listeners could hear one of two ambiguous pitches in the vicinity of the original pitch. The pitches estimated from the population-interval distributions for these respective cases (Fig. 5E) closely correspond to the pitch shifts that have been observed for human listeners. When the complex is harmonic (n=Fc/Fm=6=integer), there is one clear pitch and one population-interval maximum. When the complex is inharmonic (n=5.86=noninteger), the pitch shifts, as does this maximum (arrow). When the components are further shifted downward (n=5.5), either of the two pitches can be heard with roughly equal probability; correspondingly, two equal population-interval maxima are present (double arrows). Thus, a complex set of harmonically-based pitch effects can be readily explained in terms of population-interval representations.

The relative insensitivity of most auditory perception to the phase spectra of stationary sounds has long been recognized. For complex tones consisting of lower-frequency stimulus components (<1500 Hz), distinguishing stimuli that differ in phase, but not magnitude spectrum, despite very obvious differences in their waveforms, is generally very difficult. The waveforms of two such stimuli, an AM tone and a quasi-frequency modulated (QFM) tone that differ only in the phases of their center components (640 Hz), are shown in Fig. 5F. Their waveform envelopes are considerably different, with the AM tone having an envelope that is highly modulated, and the QFM tone having one that is much flatter. Their perceptual indistinguishability argues against auditory mechanisms that are sensitive to the phases of low-frequency components, such as neural computations that carry out an analysis of whole-waveform envelopes. The respective population-interval distributions for these stimuli are highly similar, almost indistinguishable.

Whereas the perception of pitches created by low-frequency harmonics is largely phase-insensitive, the same cannot be said for the higher-frequency, closely-spaced, perceptually "unresolved" harmonics. Alterations in the phases of the upper harmonics can affect the low pitches that they evoke (for example, doubling the pitch) by altering the shapes of waveform envelopes that are produced by cochlear filtering. Thus psychophysically, two kinds of low pitches appear to be evoked by complex tones:

1) phase-insensitive pitches that are produced by lower-frequency, perceptually-resolved harmonics, and
2) phase-sensitive pitches that are produced by higher-frequency, unresolved harmonics.

This dichotomy has led some auditory theorists to posit dual pitch mechanisms, one for resolved harmonics alongside another for unresolved ones (Carlyon & Shackleton, 1994; Shackleton & Carlyon, 1994). Both sets of low pitches can be explained, however, in terms of a central analysis of all-order population-interval distributions. For closely spaced, unresolved, higher-frequency components, phase-locking to individual components is weak relative to phase-locking to envelopes, such that the interspike intervals that are produced (primarily by high CF fibers) mostly reflect the temporal structure of the envelope. The two kinds of low pitches may therefore correspond to the two modes by which pitch-related interspike intervals can be generated: phase-locking to individual harmonics, and phase-locking to their interactions (envelopes). As population-interval distributions incorporate intervals that are generated by both mechanisms, these representations provide a unified analytical framework that accounts for both kinds of pitches (Cariani & Delgutte, 1996a, 1996b; Meddis & O'Mard, 1997). The perceptual resolvability of harmonics itself may have a neural basis in the two competing mechanisms of interval generation and in the discriminability of multiple interval peaks in population-interval distributions that they produce (cf. discussion of neural coding and

signal detection (Moore, 1997, pp. 118–121). Thus, in population-interval accounts, linkages could exist between the perceptual resolvability of harmonics and different modes of pitch perception.

The dominance of lower-frequency harmonics in determining the low pitch of a complex tone ("the dominance region for pitch") can also be explained in population-interval terms. All other factors being equal, when two harmonic complexes, one consisting of lower-frequency (<1500 Hz) and the other of higher-frequency (>1500 Hz) components, each having slightly different fundamentals, are presented together so that their pitches compete, the pitch of the former is almost always heard over that of the latter. Population-interval distributions for such a combination of two complexes (harmonics 3-5 of fundamentals at 160, 240, 320, and 480 Hz versus harmonics 6-12 of fundamentals 20% higher) are shown in (Fig. 5G). In all cases, the predominant interval in the distribution corresponds to the dominant pitch of the lower-frequency complex ($1/F0_{3-5}$) rather than that of the upper ($1/F0_{6-12}$). It thus appears that harmonics in the 500 to 1500 Hz range are disproportionately effective in generating many all-order interspike intervals at the fundamental period. These frequencies produce the most highly phase-locked responses in the greatest number of fibers. As a result, because of the basic factors that are common to many mammalian auditory systems, population-interval distributions preferentially reflect the stimulus frequencies that play a predominant role in determining pitch percepts.

In addition to pitch, vowel quality or timbre can also be represented in population-interval distributions in patterns of short intervals (Fig. 5H). Timbre itself is a complex, multidimensional auditory quality that can depend upon many factors, such as spectral shape, onset and offset properties, ongoing temporal dynamics (vibrato, roughness), and phase coherence (tones vs. noises). For stationary, harmonic sounds, timbre is determined by spectral shape, for example, the locations and heights of formants. The stimulus autocorrelation function and the population-interval distribution show a series of minor peaks, which are associated with components in the formant region that give the vowel its characteristic tone quality. Patterns of shorter intervals, those less than half the fundamental period (1/F0<6.25 ms), reflect formant structure alone, whereas patterns of longer intervals reflect fundamental-formant relationships. For multiple formant vowels, the patterns of short intervals in population-interval distributions are sufficient to discriminate different vowels, using temporal information alone (Palmer, 1992; Cariani, 1995; Cariani et al., 1997). The appearance and disappearance of minor peaks in the population-interval distribution also closely follow the perceptual vowel-class boundaries that are observed psychophysically (Hirahara et al., 1996).

These findings, when taken together with those derived from populations of simulated auditory nerve fibers, suggest that many diverse aspects of pitch perception can be directly explained in terms of population-interval distributions at the level of the auditory nerve. The main conclusions can be summarized as follows.

1) First, with very few exceptions, the most common all-order interval present in the population precisely and robustly corresponds to the pitch that is heard.
2) Second, the relative proportion of pitch-related intervals amongst all others roughly corresponds to the strength of the pitch that is heard.

Many complex aspects of pitch perception can consequently be readily explained in terms of a central analysis of population-interval representations. All-order interspike intervals themselves are time durations that preserve harmonic relations between frequencies, such as the 2:1 octave ratio. If the auditory system uses representations that preserve the harmonic structure inherent in time intervals, then the perception of basic harmonic relations may be a direct consequence of the neural codes that the auditory system uses to represent and analyze sounds, rather than the product of elaborate harmonic cognitive schemas that have been built up from prior experience.

CODING OF PITCH IN THE CENTRAL AUDITORY SYSTEM

Whether such a temporal analysis is in fact implemented in the central auditory system, what form it might take, and where it might occur are issues that are presently under investigation. Previous studies of neural responses in the auditory brainstem have indicated a widespread locking of discharges to pitch-related stimulus periodicities (Greenberg & Rhode, 1987; Kim & Leonard, 1988; Kim et al., 1990; Rhode, 1995). Several populations of neurons in the three major divisions of the cochlear nucleus (anteroventral division, AVCN; posteroventral division, PVCN; dorsal division, DCN) project to more central auditory stations in the brainstem and midbrain. By virtue of the differences in the distribution of their inputs and intrinsic properties, the neurons in each population have a characteristic response pattern when driven with pure tone bursts at their characteristic frequencies (TBCF). As in the auditory nerve, harmonic complex tones that produce strong pitches at their fundamentals similarly produce many pitch-related interspike intervals. Figure 6 shows the responses to a single-formant vowel of three physiologically-characterized units (Fig. 6A to 6C) that are representative of their respective populations.

Previous studies have identified the morphological cell-types that are associated with different TBCF response patterns (Pfeiffer, 1966; Rhode et al., 1983; Young, 1984). "Primary-like" TBCF responses are produced by spherical cells in the AVCN, sustained "chopper" responses to high-frequency tone bursts are produced by multipolar cells in the PVCN, whereas "pauser" patterns consisting of an onset-pause-sustained discharge pattern are produced by fusiform cells in the dorsal division (DCN). Primarylike units, as their name implies, have responses that are name implies, have responses that are most similar to those of primary sensory neurons (auditory nerve fibers). The discharges of this primarylike unit (Fig. 6A) exhibit stimulus-driven periodicities that are associated with fundamental (12.5 ms) and formant frequency (multiples of 1.6 ms), as well as with intrinsic periodicities that are associated with the characteristic frequency of the unit (CF=400 Hz; 1/CF=2.5 ms). These intrinsic periodicities ostensibly stem from similar CF-related periodicities that are seen in auditory nerve fibers, which are in turn produced by the mechanics of the cochlea. Sustained chopper responders are so named because they fire very regularly ("chop") at their own characteristic rate when driven by high-frequency tone bursts. When these units are driven by periodic harmonic stimuli, however, their discharges almost invariably lock strongly to the fundamental and only weakly to other stimulus-periodicities, if at all (Rhode, 1998). Pauser responders manifest more complex TBCF patterns that are the product of both intrinsic membrane properties and local circuit action. Whereas these units tend to respond more weakly to periodic stimuli than do other cochlear nucleus response types, their discharges nevertheless lock to fundamentals to produce many pitch-related intervals. A general rule of thumb for these populations is that if a unit responds to a harmonic stimulus that is capable of producing a strong low pitch, the unit will either produce intervals that are related to the fundamental (extrinsic, stimulus-driven time structure) or to its characteristic frequency (stimulus-triggered, system-dependent intrinsic time structure). As intervals related to the fundamental are common to all units that are driven by a harmonic complex tone, but those related to any given characteristic frequency are not, it is all but inevitable that such pitch-related intervals predominate in these cochlear nucleus populations (for the same reasons that such intervals predominate in the auditory nerve). Thus, the population-interval representations of pitch appear to be viable at the level of the cochlear nucleus, as well as at the auditory nerve.

From all accounts, as one ascends the auditory pathway to auditory midbrain, thalamus, and cortex, the presence of pitch-related interspike-interval information becomes less apparent. One possibility is that interspike interval information is converted to a rate-based representation

somewhere in the pathway. Units that are differentially responsive to particular modulation frequencies have been proposed as the basis of such a time-to-place transformation (Langner, 1992), although whether such rate-based representations are sufficiently precise or robust to account adequately for pitch perception is not yet clear.

Another possibility is that interspike interval information persists, albeit in a sparser and more distributed form, at still more central stations. The same amount of interval information might well be distributed more sparsely over progressively greater numbers of neurons. Intervals bearing periodicity-related information might be multiplexed with other kinds of spike patterns bearing information about location and context. These factors would make interspike interval information more difficult to detect using standard spike train analysis techniques.

Still another possibility is that central stations might simply use less interval information than is available at more peripheral stations. A great overabundance of interval-based information exists in the auditory nerve, such that relatively small numbers of intervals are sufficient to account for the high precision of frequency discrimination (Siebert, 1970). Indeed, this overabundance has often been used to assert that if auditory central processors were to make

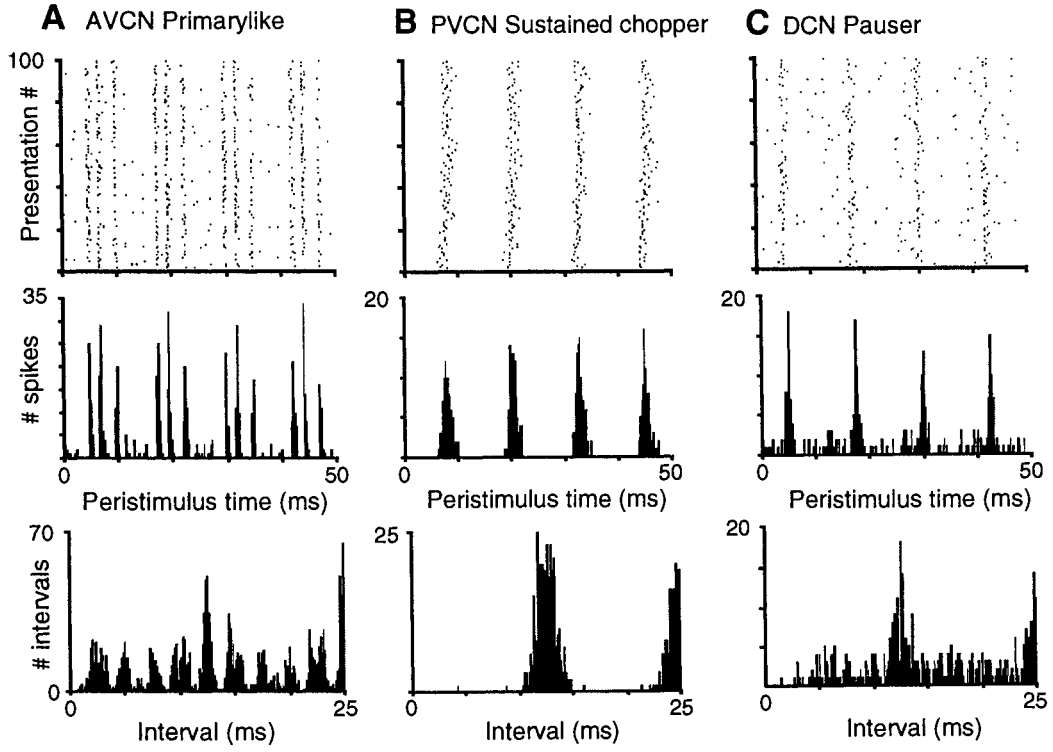

Fig. 6: Responses of three units in the cochlear nucleus to 100 presentations of a single-formant vowel (F0=80 Hz, F1=640 Hz, BW=50) at 60 dB SPL. Units were classified according to their PSTH response to short tone bursts at CF. A. Dot-raster, PSTH, and all-order interval histogram for a primarylike unit in antero-ventral cochlear nucleus (AVCN), CF=400 Hz. B. Response of a sustained chopper unit in posterior-ventral cochlear nucleus (PVCN), CF=1.5 kHz. C. Response of a pauser unit in dorsal cochlear nucleus (DCN), CF=4.4 kHz.

optimal use of this information, then human frequency discrimination would be some 40 times better than it is. The other side of this coin is that even if most interval information were to be lost or degraded in the ascending pathway, then enough information would remain to account for the observed precision of pitch discrimination.

Although stimulus-driven temporal structure declines at higher stations, it is important not to understate how much remains. Most studies thus far have been conducted under general anesthesia, but such agents generally reduce the upper frequency limits of stimulus-driven neural response periodicities by about one half (for example, see Goldstein et al., 1959). In unanesthetized animals, considerable phase-locking to 1–2 kHz tones is observed at the thalamic level (de Ribaupierre, 1997). Likewise, in the input layers of unanesthetized primary auditory cortex, fundamental frequencies up to 400 Hz are reflected in the synchronized responses of local ensembles of auditory neurons (Steinschneider et al., 1998). First-spike latencies for onsets of tone bursts at the level of the primary auditory cortex have small variances on the order of fractions of a millisecond that are comparable to those seen at the auditory nerve (Phillips, 1989; Heil, 1997), despite a conspicuous lack of sustained phase-locking to the pure tones themselves. Precise temporal patterns, embedded in spike trains, occur in a diversity of cortical locations (Abeles et al., 1993; Lestienne & Tuckwell, 1998). Although the evidence for and against time structure in the cerebral cortex has a decidedly mixed character, the data nevertheless suggest that the cortex may be capable of preserving more fine timing information than is commonly thought.

GENERAL IMPLICATIONS FOR SENSORY CODING

Population-interval representations hold many general implications for sensory coding. The coding of pitch through the interspike interval statistics of a population of neurons is a strong example of a temporal pattern code in its purest form, an example of a distributed temporal population code that does not entail interneural synchrony.

The population-interval distribution differs from both rate- and channel-based representations in two crucial ways: through the different nature of their primitives, and through the qualitatively different roles that channels play. Interspike intervals are time intervals that describe temporal relations between pairs of jointly occurring spike events. Such time intervals constitute correlational, relational primitives. In contrast, representations that are based on probabilities or on rates of unitary spike events count numbers of spike events over contiguous time windows. The counting assumption, with its scalar signals, in turn necessitates a whole host of assumptions concerning the functional topology of neural networks (Cariani, 1997). Second, in a population-interval representation, channel identities, which particular channels are activated how much, are not essential to the representational function. In the auditory nerve, of course, particular CF regions are preferentially activated by stimulus components that are nearby in frequency, and these regions will therefore contribute relatively more of their stimulus-related intervals to the global distribution. In this way, the population-interval distribution reflects the differential contributions of different CF regions. Once the intervals are combined, however, the representation does not rely on the particular channel-identities of the fibers to encode frequency (because the intervals themselves bear this information, and in a much more precise and robust way). One could discard all information concerning characteristic frequency (or cochlear place) without affecting the representation. In contrast, in a channel-based neural representation, such as a rate-place frequency map, the identities of particular channels are absolutely critical for representational function. Consequently, stimulus representations would be corrupted if the channel-identities were scrambled (if the "labels" on the

"labeled lines" were switched). Thus, the population-interval representation relies upon how neurons in a population respond and which intervals they produce, rather than upon which particular neurons fire how much.

The basic informational constituents of rate-place and population-interval representations are therefore very different, such that they complement each other, with neither representational mode precluding the other. The same holds true for representations based on relative response latencies and neural synchronies: fine temporal structure and relative latency patterns can all coexist within the coarser-grained, tonotopically ordered spatial patterns of activation. At the level of the auditory nerve and cochlear nucleus, the representation of periodicity pitch appears to follow this pattern of fine temporal structure within more coarsely tuned frequency channels.

Strong correspondences, population-interval distributions, and their respective stimulus autocorrelation functions were manifest in the similar patterns of major and minor peaks, with major peaks corresponding to pitch and minor peaks to timbres. Pitch judgments are relatively well-described by temporal autocorrelation models, precisely because the neural, interval-based representations subserving these judgments are themselves autocorrelation-like. In retrospect, the reasons for such similarities are fairly straightforward, being direct consequences of the stimulus-locked nature of auditory nerve fiber discharges. In the cochlea, the acoustic stimulus is, in effect, passed through a set of band-pass filters, such that each auditory nerve fiber is driven by a different set of frequency components. If a signal is passed through an array of overlapping frequency channels consisting of linear band-pass filters, then a series of filtered waveforms is produced. The sum of the channel-autocorrelations of the filtered waveforms equals the autocorrelation of the original, unfiltered signal (Licklider, 1951) for the same formal reasons that permit the linear superposition of Fourier components in power spectra. In the cochlea, each auditory nerve fiber produces phase-locked discharges to components whose frequencies are closest to its own CF, and in doing so, produces all-order intervals that are correlated with the autocorrelations of those components. The interval peak positions for individual fibers consequently mirror those in the stimulus autocorrelation (Fig. 4). In a linear system, both peak positions and relative heights would mirror those in the stimulus autocorrelation function. Whereas the population-interval distributions presented in Fig. 3F show similar peak positions, the relative peak heights are noticeably different. Such differences are created by nonlinear processes, such as firing rate thresholds and saturations, that alter relative peak heights without changing peak positions. Representations of frequency that are based on all-order interspike intervals are therefore resistant to many kinds of intensity-dependent nonlinearities. The functional implications of nonlinear distortions in the cochlea thus depend critically on the neural codes that the central auditory system uses. The robustness of interval-based representations, with respect to intensity-dependent distortions, makes them ideal for representing auditory forms.

More generally, it can be said that to the extent that spike-arrival times are correlated with a stimulus waveform, the intervals between spikes will be correlated with the stimulus autocorrelation function. This relation will hold for any sensory system whose receptors follow the time courses of their effective stimuli. Such phase-locking is seen for patterns of vibrations on the skin (Morley et al., 1990; Mountcastle, 1993) and for changing luminance patterns as images move relative to retinal arrays (Reichardt, 1961; Pollen et al., 1989; Bialek et al., 1991). Autocorrelation-like sensory representations that use all-order interspike intervals thus constitute potential stimulus-coding strategies in such modalities. Representations of visual form and texture that are based on spatial autocorrelation have been proposed (Uttal, 1975), but few attempts have been made to use stimulus-driven, fine spatiotemporal correlation structure for this purpose (Reitboeck et al., 1988). Recent psychophysical evidence points to a strong role for such

structure in the perception of visual forms (Lee & Blake, 1999).

The means by which neural computational architectures might make use of interspike interval statistics of populations of neurons largely remains to be explored. Different kinds of codes naturally lead to different neural-processing architectures. Channel-coding naturally leads to connectionist networks, in which information is represented through specific patterns of channel activation and processed through networks, in which specific connectivities determine functional roles. In such systems, highly specific modification of effective connectivity is the main mechanism by which functional plasticity is achieved.

For the most part, when temporal structure has been considered in functional terms; it has been assumed that to use the information, temporal patterns must be converted to channel-activation patterns. Thus the first neural auditory computation networks converted time-of-arrival differences and temporal patterns into spatial patterns of activations. Time-delay neural architectures, consisting of tapped delay lines and coincidence counters, were proposed for using interaural time-of-arrival differences to localize sounds by computing binaural cross-correlations (Jeffress, 1948). The coincidence channels that were maximally activated served to indicate the relative time-of-arrival of sounds at the two ears and hence, their location in the azimuthal plane. Similarly, neural time-delay networks that used a different arrangement of tapped delay lines and coincidence counters were proposed for carrying out neural autocorrelational analyses that spatialize population-interval distributions for analysis (Licklider, 1951; Lyon & Shamma, 1995). The coincidence counters act as autocorrelating periodicity-detectors that operate on all-order intervals.

In general, systematic differences in arrival times of external disturbances at different sensory surfaces support neural representations for external location based on stimulus-locked, time-of-arrival codes. Such differences naturally lend themselves to analysis via temporal cross-correlation operations (Carr, 1993). On the other hand, characteristic temporal spike-patterns that are generated at sensory surfaces through either stimulus-locking or stimulus-triggered intrinsic responses potentially support neural representations of stimulus-form. These characteristic temporal patterns naturally lend themselves to autocorrelational analyses. Early comprehensive computational models for hearing (Licklider, 1959; Cherry, 1961) integrated both kinds of correlational processes to represent both location and form. How many auditory functions can be subsumed under these two operations, and the extent to which other sensory systems might operate using similar principles, remains to be seen.

In time-delay architectures, plasticity of function is achieved by adjusting effective connectivities to favor particular sets of time-delays (Licklider, 1959; Tank & Hopfield, 1987), or by adjusting time delays to synchronize particular sets of inputs (MacKay, 1962). Changes in temporal response properties as a result of conditioning have been observed in a wide variety of systems (Morrell, 1967; Thatcher & John, 1977; John & Schwartz, 1978; Singer, 1995). In principle, a neural assembly can be formed that will respond preferentially to any spatiotemporal pattern in its inputs by adjusting the relative time delays and connection weights to match those in the incoming pattern. Neural delays can be created by any process that takes time to unfold and be modified by any process that alters response latency. Axonal and dendritic transmission times, latencies of activation, and time-courses of neuronal recovery (Raymond, 1979; Wasserman, 1992) potentially provide shifts in time and sensitivities to time pattern that can become control points for adaptive adjustment. Intraneural delays can then be concatenated in multisynaptic, recurrent, and/or re-entrant pathways to form still longer delays. To the extent that timing is important in a neural information processing system, such alterations of temporal response properties provide avenues by which modifications of structure can lead to modifications of function.

Finally, neural networks that carry out their operations entirely within the time domain can be

envisioned (Cariani, in press). Neural timing nets, consisting of tapped delay lines and coincidence detectors, analyze temporally-coded inputs to produce temporally-coded outputs. Simple feed-forward timing networks function as temporal sieves that extract common periodicities in their inputs, thereby finding similarities and differences between them. A fundamental advantage of these timing nets is that they operate on interval statistics, obviating the necessity for precise regulation of point-to-point connectivities. Recurrent timing networks can be used to build up periodic temporal patterns in their inputs and to separate out repeating patterns that have different periods. Combinations of feed-forward and recurrent delay lines coupled with coincidence and anticoincidence elements may then provide general-purpose strategies for detecting correlational, relational structure in the world. Efforts to understand the combined functional capabilities of temporal codes and timing nets are presently in their early, formative stages.

ACKNOWLEDGMENTS

This work was supported by Grant DC03054 from the National Institute for Deafness and Communications Disorders (NIDCD) of the National Institutes of Health (NIH).

REFERENCES

Abeles M, Bergman H, Margalit E, Vaadia E. Spatio-temporal firing patterns in the frontal cortex of behaving monkeys. J. Neurophysiol. 1993; 70: 1629–1638.

Barlow HB. The neuron doctrine in perception. In: Gazzaniga, MS, ed, The Cognitive Neurosciences. Cambridge: MIT Press 1995; 415–435.

Bialek W, Rieke F, van Stevenink RR, de Ruyter WD. Reading a neural code. Science 1991; 252: 1854–1856.

Boring EG. Sensation and Perception in the History of Experimental Psychology. New York: Appleton-Century-Crofts, 1942.

Bower TGR. The evolution of sensory systems. In: MacLeod RB, Pick Jr. H, eds, Perception: Essays in Honor of James J. Gibson. Ithaca, NY: Cornell University Press 1974; 141–152.

Brugge JF, Reale RA, Hind JE. The structure of spatial receptive fields of neurons in primary auditory cortex of the cat. J Neurosci 1996; 16: 4420–4437.

Cariani P. As if time really mattered: temporal strategies for neural coding of sensory information. Communication and Cognition—Artificial Intelligence (CC-AI) 1995; 12: 161–229. Reprinted in: Pribram K, ed, Origins: Brain and Self-Organization. Hillsdale, NJ: Lawrence Erlbaum 1994; 208–252.

Cariani P. Emergence of new signal-primitives in neural networks. Intellectica 1997; 1997: 95–143.

Cariani P. Temporal coding of sensory information. In: Bower, JM, ed, Computational Neuroscience: Trends in Research. 1997. New York: Plenum 1997; 591–598.

Cariani P. Neural timing nets for auditory computation. In: Greenberg S, Slaney M, eds, Computational Models of Auditory Function. Amsterdam: IOS Press in press; 16 pp.

Cariani P. Delgutte B. Tramo M. Neural representation of pitch through autocorrelation. Proceedings. Audio Engineering Society Meeting (AES). New York. September. 1997. Preprint #4583 (L-3); 1997.

Cariani PA, Delgutte B. Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. J Neurophysiol. 1996a; 76: 1698–1716.

Cariani PA, Delgutte B. Neural correlates of the pitch of complex tones. II. Pitch shift. pitch ambiguity. phase-invariance. pitch circularity. and the dominance region for pitch. J Neurophysiol. 1996b; 76: 1717–1734.

Carlyon RP. Shackleton TM. Comparing the fundamental frequencies of resolved and unresolved harmonics: evidence for two pitch mechanisms?. J Acoust. Soc. Am. 1994; 95: 3541–3554.

Carr CE. Processing of temporal information in the brain. Ann Rev Neurosci 1993; 16: 223-243.

Cherry C. Two ears—but one world. In: Rosenblith WA, ed, Sensory Communication. New York: MIT Press/ John Wiley 1961; 99–117.

Chung SH. Raymond SA. Lettvin JY. Multiple meaning in single visual units. Brain Behav Evol 1970; 3: 72–101.

Covey E. Temporal Neural Coding in Gustation. Duke University. 1980.

de Boer E. On the "residue" and auditory pitch perception. In: Keidel WD, Neff WD, eds, Hand-book of Sensory Physiology. Berlin: Springer Verlag 1976; 479–583.

de Cheveigné A. A pitch perception model. ICASSP 86 (Tokyo) 1986; 897–900.

de Ribaupierre F. Acoustical information processing in the auditory thalamus and cerebral cortex. In: Ehret G, Romand R, eds, The Central Auditory System. New York: Oxford University Press 1997; 317–397.

Deadyler SA, Hampson RE. Ensemble activity and behavior: what's the code?. Science 1995; 270: 1316–1318.

Delgutte B. Physiological models for basic auditory percepts. In: Hawkins H, McMullin T, Popper AN, Fay RR, eds, Auditory Computation. New York: Springer Verlag, 1995, 157–220.

Di Lorenzo PM, Hecht GS. Perceptual consequences of electrical stimulation in the gustatory system. Behavioral Neuroscience 1993; 107: 130–138.

Emmers R. Pain: A Spike-Interval Coded Message in the Brain. New York: Raven Press 1981.

Evans EF. Place and time coding of frequency in the peripheral auditory system: some physiological pros and cons. Audiology 1978; 17: 369–420.

Gerstner W. Spiking neurons. In: Maass W, Bishop CM, eds, Pulsed Neural Networks. Cambridge, Massachusetts, USA: MIT Press 1999; xiii–xxvi.

Gesteland RC, Lettvin JY, Pitts WH. Chemical transmission in the nose of the frog. J Physiol. 1965; 181: 525–559.

Goldstein MH Jr, Kiang NYS, Brown RM. Responses of the auditory cortex to repetitive acoustic stimuli. J Acoust Soc Am 1959; 31: 356–364.

Greenberg S, Rhode WS. Periodicity coding in cochlear nerve and ventral cochlear nucleus. In: Yost WA, Watson CS, eds, Auditory Processing of Complex Sounds. Hillsdale, NJ: Lawrence Erlbaum Associates 1987; 225–236.

Hatsopoulos NG, Ojakangas CL, Donohue JP, Maynard EM. Detection and identification of ensemble codes in motor cortex. In: Eichenbaum HB, Davis JL, eds, Neuronal Ensembles: Strategies for Recording and Decoding. New York: Wiley-Liss 1998; 161–176.

Heil P. First-spike timing of auditory-nerve fibers and comparison with auditory cortex. J Neurophysiol. 1997; 78: 2438–2454.

Hirahara T, Cariani P, Delgutte B. Representation of low-frequency vowel formants in the auditory nerve. Proceedings. European Speech Communication Association (ESCA) Research Workshop on The Auditory Basis of Speech Perception. Keele University, UK, July 15–19, 1996; 1–4.

Jeffress LA. A place theory of sound localization. J Comp Physiol Psychol 1948; 41: 35–39.

John ER. Switchboard vs. statistical theories of learning and memory. Science 1972; 177: 850–864.

John ER. Representation of information in the brain. In: John ER, ed, Machinery of the Mind. Boston, MA: Birkhauser 1990; 27–56.

John ER, Schwartz EL. The neurophysiology of information processing and cognition. Ann Rev Psychol 1978; 29: 1–29.

Kauer JS. Response patterns of amphibian olfactory bulb neurones to odour stimulation. J Physiol 1974; 243: 695–715.

Kiang NYS. Stimulus representation in the discharge patterns of auditory neurons. In: Tower DB, ed, The Nervous System. New York: Raven Press 1975; 81–96.

Kiang NYS. Moxon EC. Tails of tuning curves of auditory-nerve fibers. J Acoust Soc Am 1974; 55: 620–630.

Kiang NYS, Watanabe T, Thomas EC. Clark LF. Discharge Patterns of Single Fibers in the Cat's Auditory Nerve. Cambridge, Massachusetts, USA: MIT Press 1965.

Kim DO, Leonard G. Pitch-period following response of cat cochlear nucleus neurons to speech sounds. In: Duifhuis H, Horst JW, Wit HP, eds, Basic Issues in Hearing. London: Academic Press 1988; 252–260.

Kim DO, Molnar CE. A population study of cochlear nerve fibers: comparison of spatial distributions of average-rate and phase-locking measures of responses to single tones. J Neurophysiol. 1979; 42: 16–30.

Kim DO, Sirianni JG, Chang SO. Responses of DCN-PVCN neurons and auditory nerve fibers in unanesthetized decerebrate cats to AM and pure tones: Analysis with autocorrelation/power-spectrum. Hearing Res 1990; 45: 95–113.

Kozak WM, Reitboeck HJ. Color-dependent distribution of spikes in single optic tract fibers of the cat. Vision Res 1974; 14: 405–419.

Langner G. Periodicity coding in the auditory system. Hearing Res 1992; 60: 115–142.

Laurent G, Naraghi M. Odorant-induced oscillations in mushroom bodies of the locust. J Neurosci. 1994; 14: 2993–3004.

Lee D, Port NL, Kruse W, Georgoupoulos AP. Neuronal population coding: multi-electrode recordings in primate cerebral cortex. In: Eichenbaum HB, Davis JL, eds, Neuronal Ensembles: Strategies for Re-cording and Decoding. New York: Wiley-Liss 1998; 117–138.

Lee S-H, Blake R. Visual form created solely from temporal structure. Science 1999; 284: 1165–1168.

Lestienne R, Strehler BL. Time structure and stimulus dependence of precise replicating patterns present in monkey cortical neuronal spike trains. Brain Res 1987; 43: 214–238.

Lestienne R, Tuckwell HC. The significance of precisely replicating patterns in mammalian CNS spike trains. Neuroscience 1998; 82: 315–336.

Licklider JCR. A duplex theory of pitch perception. Experientia 1951; VII: 128–134.

Licklider JCR. Three auditory theories. In: Koch S, ed, Psychology: A Study of a Science. Study I. Conceptual and Systematic. New York: McGraw-Hill 1959; 41–144.

Lyon R, Shamma S. Auditory representations of timbre and pitch. In: Hawkins H, McMullin T, Popper AN, Fay RR, eds, Auditory Computation. New York: Springer Verlag 1995; 221–270.

MacKay DM. Self-organization in the time domain. In: Yovitts, MC, Jacobi GT, Goldstein GD., eds, Self-Organizing Systems 1962. Washington. DC: Spartan Books 1962; 37–48.

Macrides F. Dynamic aspects of central olfactory

processing. In: Schwartze DM, Mozell MM, eds, Chemical Signals in Vertebrates. New York: Plenum 1977; 207–229.

Macrides F, Chorover SL. Olfactory bulb units: activity correlated with inhalation cycles and odor quality. Science 1972; 175: 84–86.

Marion-Poll F, Tobin TR. Temporal coding of pheromone pulses and trains in Manduca sexta. J Comp Physiol A 1992; 171: 505–512.

Meddis R. Hewitt MJ Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I. Pitch identification. J Acoust Soc Am 1991; 89: 2866–2882.

Meddis R, O'Mard L. A unitary model of pitch perception. J Acoust. Soc. Am. 1997; 102: 1811–1820.

Moore BCJ. Introduction to the Psychology of Hearing, 4th Ed. London: Academic Press 1997; 118-121.

Morley JW, Archer JS, Ferrington DG, Rowe MJ, Turman AB. Neural coding of complex tactile vibration. Information Processing in Mammalian Auditory and Tactile Systems. New York: Alan R. Liss 1990; 127–140.

Morrell F. Electrical signs of sensory coding. In: Quarton GC, Melnechuck T, Schmitt FO, eds, The Neurosciences: A Study Program. New York: Rockefeller University Press 1967; 452–469.

Mountcastle V. The problem of sensing and the neural coding of sensory events. In: Quarton GC, Melnechuk T, Schmitt FO, eds, The Neurosciences: A Study Program. New York: Rockefeller University Press 1967; 393–497.

Mountcastle V. Temporal order determinants in a somatosthetic frequency discrimination: sequential order coding. Annals New York Acad Sci 1993; 682: 151–170.

Onoda N, Mori K. Depth distribution of temporal firing patterns in olfactory bulb related to air intake cycles. J Neurophysiol 1980; 44: 29–39.

Palmer AR. Segregation of the responses to paired vowels in the auditory nerve of the guinea pig using autocorrelation. In: Schouten MEH, ed, The Auditory Processing of Speech. Berlin: Mouton de Gruyter 1992; 115–124.

Perkell DH, Bullock TH. Neural Coding. Neurosciences Research Program Bulletin 1968; 6: 221–348.

Pfeiffer RR. Classification of response patterns of spike discharges from units in the cochlear nucleus: tone-burst stimulation. Exp Brain Res 1966; 1: 220–235.

Phillips DP. Timing of spike discharges in cat auditory cortex neurons: implications for encoding of stimulus periodicity. Hearing Res 1989; 40: 137–146.

Pollen DA, Gaska JP, Jacobson LD. Physiological constraints on models of visual cortical function. In: Cotterill RMJ, ed, Models of Brain Function. Cambridge, UK: Cambridge University Press 1989; 115–136.

Raymond SA. Effects of nerve impulses on threshold of frog sciatic nerve fibres. J Physiol (Lond) 1979; 290: 273–303.

Reichardt W. Autocorrelation. a principle for the evaluation of sensory information by the central nervous system. In: Rosenblith WA, ed, Sensory Communication. New York: MIT Press/John Wiley 1961; 303–317.

Reitboeck HJ Pabst M. Eckhorn R. Texture description in the time domain. In: Cotterill RMJ, ed, Computer Simulation in Brain Science. Cambridge, UK, Cambridge University Press 1988.

Rhode WS. Interspike intervals as correlates of periodicity pitch in cat cochlear nucleus. J Acoust Soc Am 1995; 97: 2414–2429.

Rhode WS. Neural encoding of single-formant stimuli in ventral cochlear nucleus of the chinchilla. Hearing Res 1998; 117: 39–56.

Rhode WS, Smith PH, Oertel D. Physiological response properties of cells labeled intracellularly with horseradish peroxidase in cat ventral cochlear nucleus. J Comp Neurol 1983; 213: 448–463.

Richmond BJ, Gawne TJ. The relationship between neuronal codes and cortical organization. In: Eichenbaum HB, Davis JL, eds, Neuronal Ensembles: Strategies for Recording and Decoding. New York: Wiley-Liss 1998; 57–80.

Richmond BJ, Optican LM, Podell M, Spitzer H. Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. I. Response characteristics.. J Neurophysiol 1987; 57: 132–146.

Rieke F, Warland D, de Ruyter van Steveninck R, Bialek W. Spikes: Exploring the Neural Code. Cambridge. MA, USA: MIT Press 1997; 395.

Rose JE. Neural correlates of some psychoacoustical experiences. In: McFadden D, ed, Neural Mechanisms of Behavior. New York: Springer Verlag 1980; 1–33.

Rose JE, Brugge JR, Anderson DJ, Hind JE. Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey. J Neurophysiol 1967; 30: 769–793.

Ryugo DK. The auditory nerve: peripheral innervation. cell body morphology. and central projections. In: Webster DB, Popper AN, Fay RR, eds, The Mammalian Auditory Pathway: Neuroanatomy. New York: Springer-Verlag 1992; 23–65.

Shackleton TM, Carlyon RP. The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination. J Acoust Soc Am 1994; 95: 3529–3540.

Shadlen MN, Newsome WT. Noise. neural codes and cortical organization. Curr Op Neurobiol 1994; 4: 569–579.

Shadlen MN, Newsome WT. The variable discharge of cortical neurons: implications for connectivity. Computation. and information coding. J Neurosci 1998; 18: 3870–3896.

Siebert WM. Stimulus transformations in the peripheral auditory system. In: Kollers PA, Eden M, eds, Recognizing Patterns. Cambridge, Massachusetts, USA: MIT Press 1968; 104–133.

Siebert WM. Frequency discrimination in the auditory system: place or periodicity mechanisms? Proc IEEE 1970; 58: 723–730.

Singer W. Search for coherence: a basic principle of cortical self-organization. Concepts Neurosci 1990; 1: 1–26.

Singer W. Development and plasticity of cortical processing architectures. Science 1995; 270: 758–764.

Singer W. Putative functions of temporal correlations in neocortical processing. In: Koch C, Davis CL, eds, Large-Scale Neuronal Theories of the Brain. Cambridge, MA, USA: MIT Press 1995; 201–237.

Slaney M, Lyon RF. On the importance of time—a temporal representation of sound. In: Cooke M, Beet S, Crawford M, eds, Visual Representations of Speech Signals. New York: Wiley 1993; 95–118.

Srulovicz P, Goldstein JL. Central spectral patterns in aural signal analysis based on cochlear neural timing and frequency filtering. IEEE, Tel Aviv. Israel, 1977.

Steinschneider M. Reser DH. Fishman YI. Arezzo J. Click train encoding in primary auditory cortex of the awake monkey: Evidence for two mechanisms subserving pitch perception. J Acoust Soc Am 1998; 104: 2395–2955.

Stevens SS. Sensory power functions and neural events. In: Loewenstein WR, ed, Principles of Receptor Physiology. Berlin: Springer-Verlag 1971; 226–242.

Tank DW, Hopfield JJ Neural computation by concentrating information in time. Proc Natl Acad Sci USA 1987; 84: 1896–1900.

Thatcher RW. John ER. Functional Neuroscience. Vol. I. Foundations of Cognitive Processes. Hillsdale, New Jersey, USA: Lawrence Erlbaum 1977; 382.

Troland LT. The psychophysiology of auditory qualities and attributes. 1929; 2: 28–58.

Uttal WR. The Psychobiology of Sensory Coding. New York: Harper Row 1973.

Uttal WR. An Autocorrelation Theory of Form Detection. New York: Wiley 1975.

van Noorden L. Two channel pitch perception. In: Clynes M, ed, Music. Mind and Brain. New York: Plenum 1982; 251–269.

Wasserman GS. Isomorphism. task dependence. and the multiple meaning theory of neural coding. Biol Signals 1992; 1: 117–142.

Wever EG. Theory of Hearing. New York: Wiley 1949.

Young ED. Response characteristics of neurons in the cochlear nuclei. In: Berlin C, ed, Hearing Science. San Diego, CA, USA: College Hill 1984; 423–460.

Young RA. Some observations on temporal coding of color vision: psychophysical results. Vision Res 1977; 17: 957.

Acoust. Sci. & Tech. **22**, 2 (2001)

TUTORIAL

Temporal coding of sensory information in the brain

Peter A. Cariani

Department of Otology and Laryngology, Harvard Medical School,
Eaton Peabody Laboratory, Massachusetts Eye & Ear Infirmary, 243 Charles St., Boston, MA 02114 USA
e-mail: peter@epl.meei.harvard.edu

Abstract: Physiological and psychophysical evidence for temporal coding of sensory qualities in different modalities is considered. A space of pulse codes is outlined that includes 1) channel-codes (across-neural activation patterns), 2) temporal pattern codes (spike patterns), and 3) spike latency codes (relative spike timings). Temporal codes are codes in which spike timings (rather than spike counts) are critical to informational function. Stimulus-dependent temporal patterning of neural responses can arise extrinsically or intrinsically: through stimulus-driven temporal correlations (phase-locking), response latencies, or characteristic timecourses of activation. Phase-locking is abundant in audition, mechanoception, electroception, proprioception, and vision. In phase-locked systems, temporal differences between sensory surfaces can subserve representations of location, motion, and spatial form that can be analyzed via temporal cross-correlation operations. To phase-locking limits, patterns of all-order interspike intervals that are produced reflect stimulus autocorrelation functions that can subserve representations of form. Stimulus-dependent intrinsic temporal response structure is found in all sensory systems. Characteristic temporal patterns that may encode stimulus qualities can be found in the chemical senses, the cutaneous senses, and some aspects of vision. In some modalities (audition, gustation, color vision, mechanoception, nocioception), particular temporal patterns of electrical stimulation elicit specific sensory qualities.

Keywords: Temporal coding, Phase-locking, Autocorrelation, Interspike interval, Spike latency, Neurocomputation

PACS number: 43.10.Ln, 43.64.Bt, 43.66.Ba, 43.66.Hg, 43.66.Wv

1. GENERAL CLASSES OF NEURAL PULSE CODES

The neural coding problem in perception involves the identification of the neural correlates of sensory distinctions [1–8]. Sensory information can be encoded in patterns of neurons that respond (channel codes) or in temporal relations between spikes (temporal codes). Temporal codes can be further subdivided into time-of-arrival codes that rely on relative spike timings across neurons and temporal pattern codes that rely on internal patterns of spikes that are produced (Table 1). Here we review psychophysical and neurophysiological evidence that bears on temporal codes in perception.

Temporal codes utilize stimulus-dependent time structure in neural responses. This time structure can be produced either through phase-locking or intrinsic response characteristics. The simplest time-of-arrival code compares spike arrival times between two neurons to convey the time difference between them, irrespective of the temporal structure internal to each spike train. In contrast, temporal pattern codes utilize this internal time structure to convey information. Temporal pattern codes utilize interspike intervals [1,8], interval sequences [9], and timecourses of discharge [10,11] to convey information. Both time-of-arrival and temporal pattern codes permit multiplexing of different kinds of information [10], though time-division [12], frequency-division [13,14] and code-division [9] schemes. Some evidence for temporal coding exists in virtually every sensory modality [1–5,7,15–18]. This review will concentrate on time-of-arrival and temporal pattern codes that arise from both phase-locked and intrinsic temporal response properties of sensory systems.

2. TIME-OF-ARRIVAL CODES IN PHASE-LOCKED SENSORY SYSTEMS

A highly robust cue for stimulus direction is the temporal pattern of activation that it produces across different sensory surfaces. In audition, mechanoception, and electroception, there appear to be common mechanisms that make use of this cue to translate temporal differences into apparent location [3, 15, 16]. In all of these sys-

Table 1 General types of neural codes.

Code class	Response properties	Neural representation/analysis
Channel	Stimulus-driven responses	Characteristic across-neuron activation patterns
Time-of-arrival	Stimulus-locked responses	Temporal crosscorrelation
	Intrinsic responses	Characteristic latency-precedence relations
Temporal pattern	Stimulus-locked responses	Temporal autocorrelation Within neurons (interspike interval patterns) Across neurons (volley patterns)
	Intrinsic responses	Characteristic temporal response pattern

tems, receptors phase-lock to their respective adequate stimuli, such that the temporal structure of the stimulus is faithfully impressed on the timings of spikes produced by primary sensory neurons. By virtue of phase-locking, relative-times-of-arrival of a stimulus at different receptor sites are translated into relative spike latencies in their respective sensory pathways. These relative spike timings are in turn analyzed via neural delay lines and temporal coincidence detectors. This temporal cross-correlation operation appears to be common neurocomputational strategy for many sensory systems across a wide range of phyla.

A classic example is the localization of sounds in the azimuthal plane by means of interaural time differences (ITD). Humans are able to use interaural time differences as small as 20–30 μs to localize sounds. Wavefronts from sound sources not directly in front of an observer arrive at the two ears at different times. They subsequently produce phase-locked spikes in auditory nerve fibers whose relative timings reflect the interaural time differences. In the auditory brainstem, highly secure synapses, tapped delay lines, and neural coincidence detectors in effect implement binaural cross-correlation operations that provide a readout of interaural time delays, and consequently, of azimuth estimates [17,19,20]. Units whose discharge rates reflect tuning to particular interaural delays [21] are found at higher stations in the ascending auditory pathway. Central representations of auditory space may also utilize location-dependent temporal response patterns [22] and/or population-latency profiles [23].

A strikingly similar situation can be found in mechanoception, where relative temporal delays ($>$ 1 ms) of mechanical stimulation at different skin locations are perceived as differences in apparent location [15]. As in audition, perceived locations move toward the sensory surfaces that lead in time. Mechanoceptors phase-lock to skin deformations and hair displacements, such that their relative timings are impressed on the discharges of neurons at many stations in the somatosensory pathway [24, 25]. That other sensory systems exhibit comparable behavior suggests that the brain may have a generalized capacity to distinguish fine spike timings ($\sim$ 1 ms) [15].

In echolocation systems of bats and cetaceans, acoustic signals are emitted and their reflection patterns are observed. Time delays between emitted signals and their echoes provide information about distances and shapes of objects. In bats, relative times-of-arrival between spikes produced by cries and echoes permit precise estimates of target ranges and shapes that correspond to microsecond time differences [26,27].

Electroception involves another time-based active sensing strategy [16, 28]. Weakly electric fish produce sinusoidally-varying electrical fields around their bodies that are deformed by the presence of nearby external objects. These deformations alter the relative phases of the electric field at different body locations, which alter the relative latencies of spikes produced in afferent electroceptive pathways. As in the binaural example, these pathways have highly secure, low jitter connections, delay lines, and central coincidence detectors that permit extremely small time-of-arrival differences (here, $<$ 1 μs) to be distinguished.

Visual receptor arrays can also be considered as collections of receptor surfaces. Phase-locking to temporal modulations of luminance produced by moving spatial patterns is ubiquitous in the visual systems of animals [29, 30]. As a consequence of phase-locking, temporal correlations between spikes produced in different visual channels potentially provide a general neurocomputational basis for the representation of visual motion. In the fly visual system, different spike timings in neighboring ommatids are used for detection of motion [7,31,32]. This temporal cross-correlation mechanism permits rapid and precise motion from small numbers of incoming spikes to inform flight course corrections in as little as 30 ms [7,32].

Binocular vision involves cross-correlation of spatio-temporal patterns registered on the two retinas. Introduc-

tion of systematic time delays between the outputs of the two retinas can be produced by placing a neutral density filter over one eye that attenuates luminance and increases the spike latencies in that monocular pathway. When a horizontally moving object is viewed under such circumstances, an illusion of depth, known as the Pulfrich effect [3], is created by the internal temporal disparities produced by the different spike latencies in the two pathways. Provided that at least some spikes are temporally correlated with the movements of object edges across receptive fields [33], then a temporal cross-correlation operation can potentially handle binocular temporal disparities in a manner not unlike ITD processing in binaural hearing.

On the motor side, muscle stretch and joint position receptors phase-lock to muscle movements [34], while vestibular afferents phase-lock to head accelerations. These provide neural representation of the temporal structure of actions: the relative timings of muscle activations and their consequences.

3. TEMPORAL PATTERN CODES IN PHASE-LOCKED RECEPTOR SYSTEMS

To the extent that sensory receptors follow the time structure of their adequate stimuli, that structure is impressed on the discharges of sensory neurons. As we have seen above, comparisons of relative timings across receptors yields information about direction and movement of the stimulus. Pitch, timbre, rhythm, flutter, and tactile texture are examples of sensory qualities that depend on the internal time structure of the stimulus. Temporal patterns of spikes produced by arrays of phase-locking receptors reflect this internal structure and thus provide information about the temporal form of the stimulus.

In the auditory system, functional roles for phase-locked neural timing information in both binaural localization and frequency representation have been well appreciated [6, 35, 36]. Phase-locking of auditory nerve fibers to acoustic stimuli creates time intervals between spikes (interspike intervals) that are directly related to stimulus periodicities. Interval distributions consequently contain spectral information up to the limits of phase-locking (5–10 kHz, depending on species and statistical criterion). Interspike interval representations easily account for the high precision of pure tone frequency discrimination over a wide range of frequencies and sound pressure levels, and for its decline as phase-locking weakens at higher frequencies [7, 35, 37]. Interestingly, minimal human Weber fractions for frequency (0.2% for 1 kHz pure tones $= 20\ \mu$s difference in period) are comparable to minimal discriminable ITD differences in binaural localization ($\sim 20\ \mu$s).

The pitches of complex tones also appear to be explicable in terms of interspike intervals. Harmonic complex tones produce pitches at their fundamentals, even in the absence of any spectral energy at the fundamental itself ("the pitch of the missing fundamental"). Many different kinds of stimuli with very different spectral energy distributions can produce the same low, "periodicity" pitches [35, 38, 39]. The most comprehensive and successful neurally-based theories of pitch compute temporal autocorrelations using all-order interspike intervals (*i.e.* between both consecutive and nonconsecutive spikes). J.C.R. Licklider in 1951 outlined a time-delay neural network that operated on phase-locked auditory nerve responses to perform an autocorrelation analysis of interspike intervals in each frequency channel [40]. Three decades later related temporal models for pitch were proposed that summed together interspike intervals across the auditory nerve [35,39]. These population-based temporal pitch models have been tested in computer simulations [41] and in neurophysiological studies [8,42,43].

The major finding of these studies has been that the most common interval present in the auditory nerve at any given time almost invariably corresponds to the pitch that is heard (the few exceptions involve octave shifts). The fraction of pitch-related intervals amongst all others qualitatively corresponds to the strength of the pitch that is heard. Such purely temporal, population-interval representations also account for a number of complex and subtle pitch phenomena: pitches of complexes with "missing fundamentals", pitch equivalence classes (metamery), relative invariance of pitch and pitch salience with sound pressure level, pitches of stimuli having psychophysically-unresolved harmonics, the "non-spectral" pitches of amplitude modulated noise, as well as complex patterns of pitch shift that are heard for inharmonic stimuli. Temporal correlates of many of these effects are also seen in single neuron responses [44] and in human frequency-following responses (FFR) produced by the auditory midbrain [45].

These population-wide distributions of all-order interspike intervals form general-purpose temporal representations that resemble stimulus autocorrelation functions [8], which can serve as time-domain representations of stimulus spectrum. Purely temporal representations of spectral shape and vowel quality are therefore also possible. Auditory nerve fibers phase-lock to frequency components in formant regions, producing related intervals whose respective numbers depend on component intensities. Population-wide interval distributions represent formant patterns through distributions of short intervals (< 4 ms) that are characteristic of the different vowels that produced them [5,39,46,47],

The somatosensory analogue of auditory pitch is the sense of flutter-vibration [3,15,24] Discrimination of vibration rates of up to 1 kHz is based on interspike intervals produced through phase-locking of cutaneous afferents [25]. As in audition, complex tactile textures produce spectrally-richer temporal patterns of spikes [48,49].

Thus, in those sensory systems where spikes are locked to stimulus transients and ongoing periodicities, direct temporal representation of the form of the stimulus is possible. Visual form could potentially be encoded through time intervals between phase-locked spikes produced at different retinal locations. Recent demonstrations that visual forms can be created through short-term spatiotemporal motion correlations that contain no long-term spatial structure suggest some role for temporal correlation structure in perception of visual forms [50].

The eyes are in constant motion, drifting even during fixation, and many central visual neurons discharge with precise latency when contrast gradients (edges) cross their receptive fields [33]. Highest observed precisions of stimulus-driven spike timings in visual systems range from hundreds of microseconds to a few milliseconds [7, 32, 51, 52]. As in many other sensory systems, perceptual discriminations that seen in visual psychophysical tasks are much finer than those permitted by the relatively coarse spatial resolutions of individual receptive fields ("hyperacuity") [3, 7]. However, if the visual system were able to make use of fine spike timing, then these observed spike precisions of 1 millisecond would account for its acuity [53].

Spatial frequency is the visual analogue of acoustic frequency, with sinusoidally modulated gratings being the visual analogues of pure tones. Phase-locking to visual flickers of 50–100 Hz and higher in early stages of visual processing is not uncommon. When such gratings are drifted across receptive fields at constant velocities, their temporal frequency at any given point simply reflects their spatial frequency. Spike timings, as analyzed through post-stimulus time (PST), period, and all-order interval histograms (Fig. 1B), faithfully replicate the temporal structure of the drifting image. As in audition, temporal frequency can thus be accurately estimated from interspike interval statistics. From the distributions in Fig. 1, temporal modulation frequencies of 16 and 32 Hz were estimated to within 1–2% using smoothed peak picking and to 0.1 Hz using best cosine fit. As for harmonic complexes in audition, there is also a "missing" fundamental phenomenon in spatial vision [54] that may be explicable in terms of (spatial) autocorrelations.

Spatial autocorrelation models account for many aspects of visual shape and texture perception [55–57]. Neural autocorrelation mechanisms based on stationary spatial activation patterns [55] and on neural synchronies [56, 57] have been proposed. The latter, spatiotemporal mechanisms move towards a theory of visual form that is based on stimulus-driven, temporally correlated spatial structure.

4. INTRINSIC TEMPORAL PATTERNS FOR ENCODING SENSORY QUALITY

Some sensory qualities, such as distinctions of smell, taste, pain, color, and temperature do not have natural correlates in the internal structure of stimulus waveforms. In sensory systems that subserve such distinctions, there may nevertheless be intrinsic, temporal patterns of neural response that are stimulus-dependent. All sensory receptors and neural populations have impulse- and step-response patterns that reflect timecourses of underlying biophysical, biochemical, and neural processes. Classic examples of such intrinsic response patterns are the different types of tone burst responses (chopping, pausing, buildup, onset) that are seen in the cochlear nucleus.

Early lateral inhibitory interactions also set up temporal precedence relations that can create different relative response latencies for different stimuli. Such across-neuron latency patterns can convey information concerning stimulus quality and intensity [2,58,59]. Latency differences can be further magnified by feedforward lateral inhibition [60].

The most striking evidence for temporal coding in the chemical senses is found in the gustatory system. In primary gustatory neurons of the rat, neural responses exhibit tastant-specific temporal discharge patterns (Fig. 2C) [61,62]. Electrical stimulation of other rats using recorded temporal response patterns elicit orofacial expressions and behaviors that would normally be associated with the respective tastant. When these recorded, naturally-generated temporal patterns used for stimulation are scrambled, or isochronous pulse trains with the same pulse rates are used, no corresponding orofacial expressions and behaviors are produced.

In olfaction, odorant-related time patterns of neural response have been observed in a wide variety of systems [63–67] Historically, efforts to find temporal pattern primitives for smell have been confounded by concentration-dependent changes in temporal response patterns [68] and complex history-dependencies. However, phase-locking to air-intake (sniffing) cycles [69] and emergent synchronized oscillations may provide population-based reference times for latency-pattern [65, 66] and latency-offset codes [70, 71]. Artificial chemical recognition devices that analyze intrinsic temporal response patterns of optical chemosensors have been developed that outperform analyses based on averaged across-

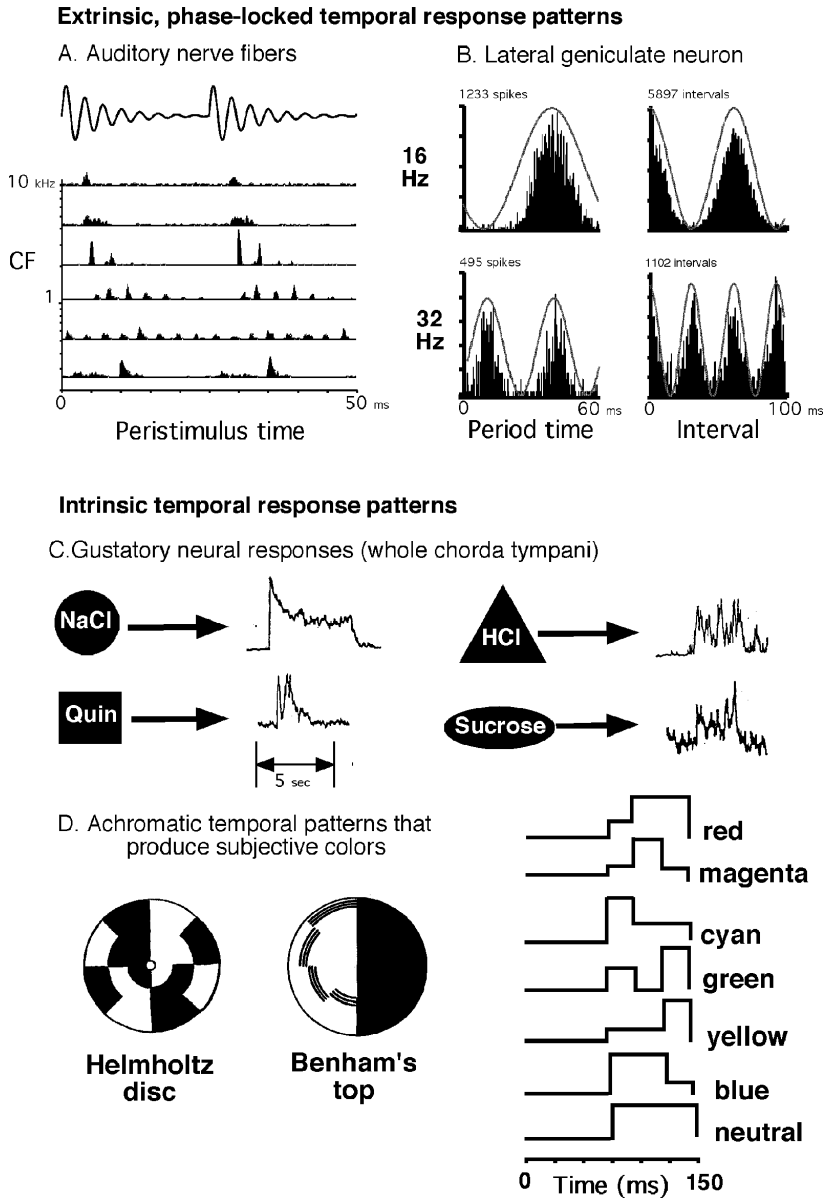

Fig. 1 Extrinsic and intrinsic temporal response patterns. A. Phase-locked responses in six auditory nerve fibers in Dial-anesthetized cat. Plot shows stimulus waveform and post-stimulus time histograms arranged by characteristic frequency (CF) for 100 presentations at 60 dB SPL. B. Phase-locked responses of a typical lateral geniculate parvo cell in anesthetized macaque to 16 and 32 Hz sinusoidal temporal modulations of luminance [83]. Period histograms and all-order interval (autocorrelation) histograms for 25 s of response. Stimulus waveform and autocorrelation have been superimposed (phase-matched). C. Intrinsic response timecourses in the gustatory system to four tastants of different classes: 0.1M NaCl (salty), 0.1M quinine (bitter), 0.1M HCl (sour), 0.5M sucrose (sweet). Waveforms are typical whole-nerve responses recorded from chorda tympani of decerebrate rats (reproduced w. permission) [61]. D. Achromatic temporal patterns that produce subjective colors. Left. Rotating patterns of Helmholtz and Benham. Right: Glow tube luminance patterns and their subjective colors (redrawn) [75].

sensor activation patterns [72].

Intrinsic temporal patterns of neural response potentially encode a number of visual attributes: texture, contrast, pattern, and color. Different patterns of visual stimulation produce different intrinsic time patterns of neural response [10,13,14,73,74].

Notable are color percepts evoked by temporal flicker. The best known demonstration of these "subjective", "achromatic" "Fechner colors" is the Benham Top (panel D), which generates temporal sequences of changing luminances and edges that correspond to particular colors [75–77]. Glow tubes have also been used to characterize the time patterns that elicit various colors [75]. Related temporal patterns of electrical stimulation of the retina produce correspondingly colored phosphenes [78]. One interpretation is that both temporally-patterned flicker and electrical stimulation extrinsically drive retinal elements to produce temporal discharge patterns that are subsequently interpreted as color signals by central visual stations. These imposed temporal patterns are presumably similar to intrinsic ones that would normally be generated by wavelength-dependent responses in the retina. Wavelength-dependent interspike interval patterns [79] latency patterns, and characteristic time courses of response [10, 14, 73] are seen in different visual systems, although their relation to subjective colors is not clear.

Historically there has been a longstanding debate over whether cutaneous sensations such as pain, touch, warmth, and cold are encoded via labeled lines (neural specificity models) or via pattern codes (channel-temporal-patterns). Interval sequence patterns have been found in thalamic units whose presence coincides with behavioral signs of pain, and whose disruption by means of appropriately patterned electrical stimulation appears to provide analgesia [9].

5. PERCEPTS EVOKED BY TEMPORALLY-PATTERNED ELECTRICAL STIMULI

The ability of a specific temporal pattern of electrical stimulation to evoke a particular sensory quality is strongly suggestive of an underlying temporal code. In a number of cases, temporally-patterned electrical stimuli produce taste [61, 62], color [78], and pain percepts [9], while randomly patterned electrical stimulation controls do not. Since electrical stimulation with gross electrodes nonspecifically drives large ensembles of neurons in the same way, it is difficult to attribute the effect to activation of particular neuronal types.

In phase-locked systems, periodic electrical stimuli produce qualities related to stimulus frequency. Periodic stimulation of the skin evokes sensations akin to flutter-vibration [25]. Electrical stimulation of the cochlea produces distinguishable pitch percepts up to roughly 500–1,000 Hz. Although auditory nerve fibers phase lock well to electrical stimuli of much higher frequencies, the highly abnormal interval patterns that are produced for periodicities above 600 Hz may explain their poor discrimination [80]. The impressive effectiveness of present day cochlear implants in restoring speech reception may be due to the critical role that lower-frequency temporal patterns (< 200 Hz) play in the speech code itself [81,82].

6. CONCLUSIONS

Stimulus-dependent temporal response structure is found in an astonishingly wide range of sensory systems. Phase-locked responses permit time differences of response to be used as cues for stimulus location that can be analyzed using neural temporal cross-correlation architectures. Phase-locking impresses stimulus time structure on neural responses, providing interval-based representations of the stimulus form that can be analyzed through neural temporal autocorrelation architectures. Intrinsic temporal response patterns produce temporal cues for stimulus properties. Psychoneural correspondences between perceptual qualities and temporal response patterns produced by natural and electrical stimuli suggest that temporal codes may subserve many more perceptual functions than is commonly believed.

ACKNOWLEDGEMENTS

This work was supported by NIH Grant DC03054 from the National Institute for Deafness and Communications Disorders, a part of the U.S. National Institutes of Health. We are most grateful to Andrzej Przybyszewski and Dan Pollen, who provided the LGN spike train data for Fig. 1B.

REFERENCES

[1] V. Mountcastle, "The problem of sensing and the neural coding of sensory events", in *The Neurosciences: A Study Program*, G. C. Quarton, T. Melnechuk and F. O. Schmitt, Eds. (Rockefeller University Press, New York, 1967), pp. 393–408.

[2] D. H. Perkell and T. H. Bullock, "Neural coding", *Neurosci. Res. Program Bull.* **6**, 221–348 (1968).

[3] W. R. Uttal, *The Psychobiology of Sensory Coding* (Harper and Row, New York, 1973).

[4] W. R. Uttal, Ed., *Sensory Coding: Selected Readings* (Little-Brown, Boston, 1972).

[5] P. Cariani, "As if time really mattered: Temporal strategies for neural coding of sensory information", *Commun. Cognit. - Artif. Intell.* (*CC-AI*) **12**, 161–229. Reprinted in *Origins: Brain and Self-Organization*, K. Pribram, Ed. (Lawrence Erlbaum, Hillsdale, N.J., 1994), pp. 208–252.

[6] E. G. Boring, *Sensation and Perception in the History of Experimental Psychology* (Appleton-Century-Crofts, New York, 1942).

[7] F. Rieke, D. Warland, R. de Ruyter van Steveninck and W. Bialek, *Spikes: Exploring the Neural Code* (MIT Press, Cam-

bridge, Mass., 1997).

[8] P. Cariani, "Temporal coding of periodicity pitch in the auditory system: An overview", *Neural Plasticity* **6**, 147–172 (1999).

[9] R. Emmers, *Pain: A Spike-Interval Coded Message in the Brain* (Raven Press, New York, 1981).

[10] B. J. Richmond, L. M. Optican and T. J. Gawne, "Neurons use multiple messages encoded in temporally modulated spike trains to represent pictures", in *Seeing Contour and Colour*, J. J. Kulikowski and C. M. Dickenson, Eds. (Pergamon Press, New York, 1989), pp. 705–713.

[11] J. D. Victor and K. Purpura, "Nature and precision of temporal coding in visual cortex", *J. Neurophysiol.* **76**, 1310–1326 (1996).

[12] W. Singer, "Time as coding space in neocortical processing", in *Temporal Coding in the Brain*, G. Buzsáki, *et al*., Eds. (Springer-Verlag, Berlin, 1994), pp. 51–80.

[13] S. H. Chung, S. A. Raymond and J. Y. Lettvin, "Multiple meaning in single visual units", *Brain Behav. Evol.* **3**, 72–101 (1970).

[14] G. S. Wasserman, "Isomorphism, task dependence, and the multiple meaning theory of neural coding", *Biol. Signals* **1**, 117–142 (1992).

[15] G. von Békésy, *Sensory Inhibition* (Princeton University Press, Princeton, 1967).

[16] C. E. Carr, "Processing of temporal information in the brain", *Annu. Rev. Neurosci.* **16**, 223–243 (1993).

[17] M. Konishi, "Deciphering the brain's codes", in *Neural Codes and Distributed Representations: Foundations of Neural Computation*, L. Abbott and T. J. Sejnowski, Eds. (MIT Press, Cambridge, Mass., 1999), pp. 1–18.

[18] J. J. Eggermont, *The Correlative Brain: Theory and Experiment in Neural Interaction* (Springer-Verlag, Berlin, 1990).

[19] C. E. Carr and M. Konishi, "A circuit for detection of interaural time differences in the brain stem of the barn owl", *J. Neurosci.* **10**, 3227–3246 (1990).

[20] S. Colburn, "Computational models of binaural processing", in *Auditory Computation*, H. Hawkins and T. McMullin, Eds. (Springer-Verlag, New York, 1996), pp. 332–400.

[21] S. Kuwada and T. C. T. Yin, "Physiological studies of directional hearing", in *Directional Hearing*, W. A. Yost and G. Gourevitch, Eds. (Springer-Verlag, New York, 1987), pp. 146–176.

[22] J. C. Middlebrooks, A. E. Clock, L. Xu and D. M. Green, "A panoramic code for sound location by cortical neurons", *Science* **264**, 842–844 (1994).

[23] J. F. Brugge, R. A. Reale and J. E. Hind, "The structure of spatial receptive fields of neurons in primary auditory cortex of the cat", *J. Neurosci.* **16**, 4420–4437 (1996).

[24] W. Keidel, "The sensory detection of vibrations", in *Foundations of Sensory Science*, W. W. Dawson and J. M. Enoch, Eds. (Springer-Verlag, Berlin, 1984), pp. 465–512.

[25] V. Mountcastle, "Temporal order determinants in a somatosthetic frequency discrimination: Sequential order coding", *Ann. NY Acad. Sci.* **682**, 151–170 (1993).

[26] J. A. Simmons, P. A. Saillant, M. J. Ferragamo, T. Haresign, S. P. Dear, J. Fritz and T. A. McMullin, "Auditory computations for biosonar target imaging in bats", in *Auditory Computation*, H. Hawkins and T. McMullin, Eds. (Springer-Verlag, New York, 1995), pp. 401–468.

[27] C. F. Moss and H.-U. Schnitzler, "Behavioral studies of auditory information processing", in *Hearing by Bats*, A. N. Popper and R. R. Fay, Eds. (Springer-Verlag, New York, 1995), pp. 87–145.

[28] W. Heiligenberg, "The coding and processing of temporal information in the electrosensory system of the fish", in *Temporal Coding in the Brain*, G. Buzaki, *et al*., Eds. (Springer-Verlag, Berlin, 1994), pp. 1–12.

[29] G. A. Horridge, "Primitive vision based on sensing change", in *Neurobiology of Sensory Systems*, R. N. Singh and N. J. Strausfeld, Eds. (Plenum Press, New York, 1989), pp. 1–16.

[30] H. Autrum, "Comparative physiology of invertebrates: Hearing and vision", in *Foundations of Sensory Science*, W. W. Dawson and J. M. Enoch, Eds. (Springer-Verlag, Berlin, 1984), pp. 1–23.

[31] W. Reichardt, "Autocorrelation, a principle for the evaluation of sensory information by the central nervous system", in *Sensory Communication*, W. A. Rosenblith, Ed. (MIT Press, Cambridge, Mass./John Wiley, New York, 1961), pp. 303–317.

[32] W. Bialek, F. Rieke, R. R. de Ruyter van Steveninck and D. Warland, "Reading a neural code", *Science* **252**, 1854–1856 (1991).

[33] F. Mechler, J. D. Victor, K. P. Purpura and R. Shapley, "Robust temporal coding of contrast by V1 neurons for transient but not for steady-state stimuli", *J. Neurosci.* **18**, 6583–6598 (1998).

[34] J. E. Rose and V. Mountcastle, "Touch and kinesis", in *Handbook of Physiology: Neurophysiology*, Vol. II, J. Field, H. W. Magoun and V. E. Hall, Eds. (American Physiological Society, Washington, D.C., 1959), pp. 387–429.

[35] B. C. J. Moore, *Introduction to the Psychology of Hearing*, 4th Ed. (Academic Press, London, 1997).

[36] J. J. Eggermont, "Functional aspects of synchrony and correlation in the auditory nervous system", *Concepts Neurosci.* **4**, 105–129 (1993).

[37] J. L. Goldstein and P. Srulovicz, "Auditory-nerve spike intervals as an adequate basis for aural frequency measurement", in *Psychophysics and Physiology of Hearing*, E. F. Evans and J. P. Wilson, Eds. (Academic Press, London, 1977), pp. 337–346.

[38] E. de Boer, "On the "residue" and auditory pitch perception", in *Handbook of Sensory Physiology*, W. D. Keidel and W. D. Neff, Eds. (Springer-Verlag, Berlin, 1976), pp. 479–583.

[39] R. Lyon and S. Shamma, "Auditory representations of timbre and pitch", in *Auditory Computation*, H. Hawkins, *et al*., Eds. (Springer-Verlag, New York, 1995), pp. 221–270.

[40] J. C. R. Licklider, "Three auditory theories", in *Psychology: A Study of a Science. Study I. Conceptual and Systematic*, S. Koch, Ed. (McGraw-Hill, New York, 1959), pp. 41–144.

[41] R. Meddis and L. O'Mard, "A unitary model of pitch perception", *J. Acoust. Soc. Am.* **102**, 1811–1820 (1997).

[42] P. A. Cariani and B. Delgutte, "Neural correlates of the pitch of complex tones. I. Pitch and pitch salience", *J. Neurophysiol.* **76**, 1698–1716 (1996).

[43] P. A. Cariani and B. Delgutte, "Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase-invariance, pitch circularity, and the dominance region for pitch", *J. Neurophysiol.* **76**, 1717–1734 (1996).

[44] G. Langner, "Periodicity coding in the auditory system", *Hear. Res.* **60**, 115–142 (1992).

[45] S. Greenberg, J. T. Marsh, W. S. Brown and J. C. Smith, "Neural temporal coding of low pitch. I. Human frequency-following responses to complex tones", *Hear. Res.* **25**, 91–114 (1987).

[46] A. R. Palmer, "Segregation of the responses to paired vowels in the auditory nerve of the guinea pig using autocorrelation", in *The Auditory Processing of Speech*, M.E.H. Schouten, Ed. (Mouton de Gruyter, Berlin, 1992), pp. 115–124.

[47] T. Hirahara, P. Cariani and B. Delgutte, "Representation of low-frequency vowel formants in the auditory nerve", in *Proceedings, European Speech Communication Association (ESCA) Research Workshop on The Auditory Basis of Speech Perception, Keele University, United Kingdom, July 15–19*, 1996 (ed. 1996), p. 4.

[48] M. Rowe, "Impulse patterning in central neurons for vibrotactile coding", in *Information Processing in Mammalian Auditory and Tactile Systems* (Alan R. Liss, Inc., New York, 1990), pp. 111–125.

[49] J. W. Morley, J. S. Archer, D. G. Ferrington, M. J. Rowe and A. B. Turman, "Neural coding of complex tactile vibration", in *Information Processing in Mammalian Auditory and Tactile Systems* (Alan R. Liss, Inc., New York, 1990), pp. 127–140.
[50] S.-H. Lee and R. Blake, "Visual form created solely from temporal structure", *Science* **284**, 1165–1168 (1999).
[51] P. Reinagel and C. Reid, "Temporal coding of visual information in the thalamus", *J. Neurosci.* **20**, 5392–5400 (2000).
[52] A. M. Zador, G. T. Buracas, M. R. DeWeese and T. D. Albright, "Efficient discrimination of temporal patterns by motion-sensitive neurons in primate visual cortex", *Neuron* **20**, 959–969 (1998).
[53] T. Carney, D. A. Silverstein and S. A. Klein, "Vernier acuity during image rotation and translation: Visual performance limits", *Vision Res.* **35**, 1951–1964 (1995).
[54] S. T. Hammett and A. T. Smith, "Temporal beats in the human visual system", *Vision Res.* **34**, 2833–2840 (1994).
[55] W. R. Uttal, *An Autocorrelation Theory of Form Detection* (Wiley, New York, 1975).
[56] M. Pabst, H. J. Reitboeck and R. Eckhorn, "A model of preattentive texture region definition based on texture analysis", in *Models of Brain Function*, R. M. J. Cotterill, Ed. (Cambridge University Press, Cambridge, England, 1989), pp. 137–150.
[57] H. J. Reitboeck, M. Pabst and R. Eckhorn, "Texture description in the time domain", in *Computer Simulation in Brain Science*, R. M. J. Cotterill, Ed. (Cambridge University Press, Cambridge, England, 1988), pp. 479–494.
[58] S. S. Stevens, "Sensory power functions and neural events", in *Principles of Receptor Physiology*, W. R. Loewenstein, Ed. (Springer-Verlag, Berlin, 1971), pp. 226–242.
[59] P. Heil, "Auditory cortical onset responses revisited: I. First-spike timing", *J. Neurophysiol.* **77**, 2616–2641 (1997).
[60] S. J. Thorpe, "Spike arrival times: A highly efficient coding scheme for neural networks", in *Parallel Processing in Neural Systems*, R. Eckmiller, G. Hartmann and G. Hauske, Eds. (Elsevier, Amsterdam, 1990), pp. 91–94.
[61] E. Covey, Temporal Neural Coding in Gustation (Ph.D. Thesis), (Duke University, 1980).
[62] P. M. Di Lorenzo and G. S. Hecht, "Perceptual consequences of electrical stimulation in the gustatory system", *Behav. Neurosci.* **107**, 130–138 (1993).
[63] G. Laurent, M. Wehr and H. Davidowitz, "Temporal representations of odors in an olfactory network", *J. Neurosci.* **16**, 3837–3847 (1996).
[64] G. Laurent, "A systems perspective on early olfactory coding", *Science* **286**, 723–728 (1999).
[65] F. Macrides and S. L. Chorover, "Olfactory bulb units: Activity correlated with inhalation cycles and odor quality", *Science* **175**, 84–86 (1972).
[66] F. Macrides, "Dynamic aspects of central olfactory processing", in *Chemical Signals in Vertebrates*, D. M. Schwartze and M. M. Mozell, Eds. (Plenum Press, New York, 1977), pp. 207–229.
[67] F. Marion-Poll and T. R. Tobin, "Temporal coding of pheromone pulses and trains in *Manduca sexta*", *J. Comp. Physiol. A* **171**, 505–512 (1992).
[68] J. S. Kauer, "Response patterns of amphibian olfactory bulb neurones to odour stimulation", *J. Physiol.* **243**, 695–715 (1974).
[69] N. Onoda and K. Mori, "Depth distribution of temporal firing patterns in olfactory bulb related to air intake cycles", *J. Neurophysiol.* **44**, 29–39 (1980).
[70] J. J. Hopfield, "Pattern recognition computation using action potential timing for stimulus representation", *Nature* **376**, 33–36 (1995).
[71] M. Wehr and G. Laurent, "Odour encoding by temporal sequences of firing in oscillating neural assemblies", *Nature* **384**, 162–166 (1996).
[72] J. White, J. S. Kauer, T. A. Dickinson and D. R. Walt, "Rapid analyte recognition in a device based on optical sensors and the olfactory system", *Anal. Chem.* **68**, 2191–2202 (1996).
[73] J. D. Victor, K. Purpura, E. Katz and B. Mao, "Population encoding of spatial frequency, orientation, and color in Macaque VI", *J. Neurophysiol.* **72**, 2151–2166 (1994).
[74] D. M. MacKay and D. A. Jeffreys, "Visually evoked potentials and visual perception in man", in *Visual Centers in the Brain*, R. Jung, Ed. (Springer-Verlag, Berlin, 1973), pp. 647–678.
[75] L. Festinger, M. R. Allyn and C. W. White, "The perception of color with achromatic stimulation", *Vision Res.* **11**, 591–612 (1971).
[76] J. J. Sheppard, *Human Color Perception: A Critical Study of the Experimental Foundation* (American Elsevier, New York, 1968).
[77] G. M. Murch, *Visual and Auditory Perception* (Bobbs-Merrill, Indianapolis, 1973).
[78] R. A. Young, "Some observations on temporal coding of color vision: Psychophysical results", *Vision Res.* **17**, 957–965 (1977).
[79] W. M. Kozak and H. J. Reitboeck, "Color-dependent distribution of spikes in single optic tract fibers of the cat", *Vision Res.* **14**, 405–419 (1974).
[80] E. Javel, "Acoustic and electrical encoding of temporal information", in *Cochlear Implants: Models of the Electrically Stimulated Ear*, J. M. Miller and F. A. Spelman, Eds. (Springer-Verlag, New York, 1988), pp. 246–295.
[81] R. V. Shannon, F. G. Zeng, V. Kamath, J. Wygonski and M. Ekelid, "Speech recognition with primarily temporal cues", *Science* **270**, 303–304 (1995).
[82] S. Rosen, "Temporal information in speech and its relevance for cochlear implants", in *Cochlear Implant: Acquisitions and Controversies*, B. Fraysse and N. Cochard, Eds. (Cochlear AG, Basel, 1989), pp. 3–26.
[83] A. Przybyszewski, J. P. Gasa, W. Foote and D. A. Pollen, "Striate cortex increases contrast gain in macaque LGN neurons", *Visual Neurosci.* **17**, 485–494 (2000).

Peter Cariani received his B.S. in Life Sciences from M.I.T. in 1978. As a graduate student he studied under theoretical biologist Howard Pattee in the Department of Systems Science at the State University of New York at Binghamton, receiving his Ph.D. in 1989. His doctoral thesis addressed evolution of new sensing functions and its implications for artificial devices. In 1990 he joined the Eaton Peabody Laboratory of Auditory Physiology as a postdoctoral fellow, and worked with Bertrand Delgutte on the coding of pitch, timbre, musical consonance, and phonetic distinctions in the auditory nerve. Dr. Cariani is currently an Assistant Professor in Otology and Laryngology at Harvard Medical School and a Research Associate at the Eaton Peabody Laboratory. He is currently developing neural timing networks that process temporally-coded patterns. With Dr. Mark Tramo, he is investigating the neural representation of pitch in the auditory cortex.

Journal of New Music Research
2001, Vol. 30, No. 2, pp. 107–135

0929-8215/01/3002-107$16.00

Temporal Codes, Timing Nets, and Music Perception

Peter Cariani

Eaton Peabody Laboratory, Massachusetts Eye and Ear Infirmary, Boston, MA, USA, and Department of Otology and Laryngology, Harvard Medical School, Boston, MA, USA

Abstract

Temporal codes and neural temporal processing architectures (neural timing nets) that potentially subserve perception of pitch and rhythm are discussed. We address 1) properties of neural interspike interval representations that may underlie basic aspects of musical tonality (e.g., octave similarities), 2) implementation of pattern-similarity comparisons between interval representations using feedforward timing nets, and 3) representation of rhythmic patterns in recurrent timing nets.

Computer simulated interval-patterns produced by harmonic complex tones whose fundamentals are related through simple ratios showed higher correlations than for more complex ratios. Similarities between interval-patterns produced by notes and chords resemble similarity-judgements made by human listeners in probe tone studies.

Feedforward timing nets extract common temporal patterns from their inputs, so as to extract common pitch irrespective of timbre and vice versa. Recurrent timing nets build up complex temporal expectations over time through repetition, providing a means of representing rhythmic patterns. They constitute alternatives to oscillators and clocks, with which they share many common functional properties.

Introduction

Music entails the temporal patterning of sound for pleasure. As such, it involves the generation of simple and complex temporal patterns and expectancies over many different time scales. Music engages both the texture of auditory qualities and the general time sense. On the shortest, millisecond time scales, periodic acoustic patterns evoke qualities of pitch and timbre, while longer patterns create rhythms and larger musical structures. How neural mechanisms in the brain subserve these perceptual qualities and cognitive structures are questions whose answers are not currently known with any degree of clarity or precision.

Music most directly engages the auditory sense. Not surprisingly, theories of music perception have developed alongside theories of auditory function, which in turn have paralleled more general conceptions of informational processes in the brain (Boring, 1942). Following Fourier, Ohm, and Helmholtz, the historically dominant view of auditory function has seen the auditory system as a running spectral analyzer. In this view, sounds are first parsed into their component frequencies by the differential filtering action of the cochlea. Filter outputs become the perceptual atoms for "central spectrum" representations, which are subsequently analyzed by central processors. In this view, neural processors that recognize harmonic patterns infer pitch, those that analyze spectral envelopes and temporal onset dynamics represent timbre, and those that handle longer, coarser temporal patterns subserve the representation of rhythm. These diverse perceptual properties are then thought to be organized into higher-order conceptual structures (images, streams, objects, schema) by subsequent cognitive processors.

An alternative view of auditory function sees time and temporal pattern as primary. While there is no doubt that the cochlea is a frequency-tuned structure, there are yet many unresolved questions as to how the brain uses patterns of

We would like to thank Mark Tramo, Marc Leman, Martin McKinney, Seth Cluett, Eric Rosenbaum, Albrecht Schneider, Malcolm Slaney, Martine Turgeon, Xaq Pitkow, and many others for useful discussions, pointers, and comments concerning the neural substrates of music perception. We would like to thank an anonymous reviewer for many useful comments and criticisms. This work was supported by DC003054 of the National Institute for Deafness and Communications Disorders of the U.S. National Institutes of Health.

Accepted: 10 July, 2001

Correspondence: Peter Cariani, Eaton Peabody Laboratory, Massachusetts Eye and Ear Infirmarry, 243 Charles St., Boston, MA 02114, USA. E-mail: cariani@mac.com. URL: www.cariani.com

cochlear and neural response to form auditory and musical percepts. A temporal-pattern theory of audition looks to temporal patterns of spikes within and across neural channels rather than spatial activation patterns amongst them. There have always existed such temporal alternatives to the frequency view: Seebeck's early acoustic demonstrations of the perceptual importance of a waveform's repetition period (de Boer, 1976), Rutherford's "telephone theory" of neural coding (Boring, 1942), the frequency-resonance theory of Troland (Troland, 1929a; Troland, 1929b), Wever's volley theory (Wever, 1949), Schouten's residue theory (Schouten, Ritsma, & Cardozo, 1962), Licklider's temporal autocorrelation model (Licklider, 1951; Licklider, 1956; Licklider, 1959), and many subsequent temporal theories of the neural coding of pitch (Cariani, 1999; Goldstein & Srulovicz, 1977; Lyon & Shamma, 1996; Meddis & O'Mard, 1997; Moore, 1997b; van Noorden, 1982). The main advantages of a temporal theory of hearing stem from the precise and robust character of temporal patterns of neural discharge. The behavior of interspike interval representations based on such discharge patterns parallels the precision and robustness of perceived auditory forms. Pitch discrimination, for example, remains precise (jnd's under 1%) over an extremely wide range of sound pressure levels (>80 dB) despite great changes that are seen in patterns of neural activation over that range. Accounting for the stability of percepts and perceptual functions over such ranges is a central problem in auditory theory that interval codes readily solve.

For music perception, a temporal theory of hearing also holds the possibility of explaining tonal and rhythmic relations in terms of the neural codes that are used to represent sound. The Pythagoreans discovered the perceptual importance of small integer ratios between frequencies (as observed through monochord lengths): the octave (2 : 1), the fifth (3 : 2), the fourth (4 : 3), major third (5 : 4) and the minor third (6 : 5). The subsequent development of the science of acoustics, running through Euclid, Galileo, Descartes, Huygens, Mersenne, Leibnitz, Euler, Rameau, D'Alembert, Saveur, Helmholtz, Mach, and many others, gradually connected these ratios with temporal vibration patterns and eventually to spatiotemporal patterns of cochlear activity (Hunt, 1978; Leman & Schneider, 1997; Mach, 1898). Existence of these special tonal relationships, which are embodied in just intonation, have always caused some music theorists to suspect that musical intervals might be rooted in innate psychological structures (DeWitt & Crowder, 1987; Hindemith, 1945; Longuet-Higgins, 1987; Schneider, 1997). Other music theorists have dismissed any special psychological role for simple ratios, in some cases on the grounds that there is no physiological basis for them in the auditory system (Mach, 1898; Parncutt, 1989). Parallel hypotheses concerning an innate neuropsychological basis for rhythmic patterns formed from simple meter ratios arise in both rhythm perception (Clarke, 1999; Epstein, 1995; Handel, 1989; Longuet-Higgins, 1987) and production (Essens & Povel, 1985; Jones, 1987).

Similarities between repeating temporal patterns whose periods are related by simple ratios are most easily appreciated in time domain representations, such as waveforms and autocorrelations. Repeating temporal patterns have inherent harmonic structure to them: patterns related by simple ratios contain common subpatterns that potentially explain the Pythagorean observations. Waveform temporal similarities are obvious, but it is entirely another step to hypothesize that the brain itself uses a time code that replicates the temporal structure of sounds. For this reason only a few attempts have been made to explicitly ground these tonal and rhythmic relations in terms of underlying temporal microstructures and neural temporal codes. A comprehensive history of the development of temporal codes and temporal microstructure in music remains to be written. It has been pointed out by an anonymous reviewer that a microtemporal basis for pitch has been proposed several times in the past, among them by the physicist Christiaan Huygens (1629–95), the Gestaltist Felix Krüger (1874–1948) and the composer and music theorist Horst-Peter Hesse (1935–present). In the field of auditory physiology, the possibility that the auditory system uses temporal pattern codes for the representation of pitch was suggested in the 1920's by L.T. Troland (Troland, 1929a; Troland, 1929b). Temporal theories of pitch were lent physiological plausibility with the work of Wever and Bray (Boring, 1942; Wever, 1949) and were lent renewed psychoacoustical plausibility with the experiments of Schouten and de Boer (de Boer, 1976). Subsequent systematic studies (Evans, 1978; Kiang, Watanabe, Thomas, & Clark, 1965; Rose, 1980) provided detailed neurophysiological grounding for later quantitative decisiontheoretic models of pure tone pitch discrimination (Delgutte, 1996; Goldstein & Srulovicz, 1977; Siebert, 1968). Unfortunately, these models rarely addressed issues, such as octave similarity, that are most relevant to the perceptual structure of pitch in musical contexts.

Perhaps the earliest explicit connection between frequency ratios and neural discharge patterns was made by J.C.R. Licklider. His "duplex" time-delay neural network (Licklider, 1951, 1956, 1959) operated on temporal discharge patterns of auditory nerve fibers to form a temporal autocorrelation representation of the stimulus. His early neurocomputational model explained a wide range of pure and complex tone pitch phenomena. Licklider (1951) states that "The octave relation, the musical third, fourth, and other consonant intervals are understandable on essentially the same [autocorrelational, neurocomputational] basis. When the frequencies of two sounds, either sinusoidal or complex, bear to each other the ratio of two small integers, their autocorrelation functions have common peaks" (p. 131).

Inspired by Licklider's theory, Boomsliter and Creel proposed their "long pattern hypothesis" for pitch, harmony, and rhythm (Boomsliter & Creel, 1962). Their harmony wheel graphically showed the temporal similarities that exist between periodic patterns related by simple ratios. They examined temporal patterns underlying musical harmony and

rhythm and postulated that the brain might process musical sounds using Licklider-style time-delay neural networks operating on different time scales.

Other auditory neurophysiologists and theorists also proposed that tonal relations and musical consonance might be grounded in the temporal firing patterns of auditory nerve fibers. The neurophysiologist Jerzy Rose, who did much seminal work on the temporal discharge patterns of auditory nerve fibers, stated that "If cadence of discharges were relevant to tone perception, one could infer that the less regular the cadence, the harsher and or rougher or more dissonant the sensory experience. If this were true, the neural data would predict a relation between consonance and frequency ratio because, in response to a complex periodic sound, the smaller the numbers in the frequency ratio the more regular is the discharge cadence. Therefore our neural data can be taken to support a frequency-ratio theory of consonance." (Rose, 1980, p. 31). On the basis of similar auditory nerve interspike interval data (Ohgushi, 1983) argued for an interspike interval basis for octave similarity. Ohgushi and others (McKinney, 1999) have also sought to explain subtle deviations from exact octave matches, the "octave stretch", in terms of interspike intervals. Roy Patterson proposed a spiral, autocorrelation-like representation based on temporal patterns of discharge that generates similar patterns when periodicities are related through small integer ratios (Patterson, 1986). This structure parallels the frequency-spiral of (Jones & Hahn, 1986). Patterson drew out a number of implications of such temporal pattern relations for musical tonality and consonance. W.D. Keidel has proposed a physiological basis for harmony in music through analysis of temporally-coded auditory signals by central neural "clock-cell" networks (Keidel, 1992; Schneider, 1997; Schneider, 2001, in press).

Over the last two decades temporal theories of pitch have evolved to incorporate population-wide interspike interval distributions, not only as specialized representations for pitch (Moore, 1997b; van Noorden, 1982), but also as more general-purpose neural representations for other auditory qualities (Cariani, 1999; Meddis & O'Mard, 1997). The implications of these global interspike interval representations for music perception are beginning to be explored. Recently Leman and Carreras (Leman, 1999; Leman & Carreras, 1997) have analyzed tonal-contextual relations between chords in a Bach piece using a perceptual module that employed a running population interval representation and a cognitive module that consisted of a Kohonen network. The neural network generates a self-organizing map of pattern similarities between the interval-based representations of the different chords, i.e., a map of chord-chord similarity relations. Their measure of similarity, the Euclidean distance in the computed map, corresponded well with related similarity judgments in human listeners (Krumhansl, 1990). More recent implementations using a spatial echoic memory (Leman, 2000) have achieved similar results without the use of a training phase.

A major problem for temporal theories of pitch has always been the nature of the central neural processors that analyze temporally-coded information. Licklider's time-delay architecture is ingenious, but such neural elements, whose action would resemble temporal autocorrelators, have not been observed at any level of the system. Pitch detectors at the level of the auditory cortex have been sought, but not found (Schwarz & Tomlinson, 1990). Time-to-place transformations could be carried out by means of modulation-tuned units that have been observed at every major station in the auditory pathway (Langner, 1992). This is the best neurally-grounded account that we currently have, but unfortunately many of the properties of the resulting central representations are highly at variance with the psychophysics of pitch. These problems are discussed more fully in later sections. This leaves auditory theory without a satisfactory central neural mechanism that explains the precision and robustness of pitch discriminations.

As a consequence of the difficulties inherent in a time-to-place transformation, we have been searching for alternative means by which temporally-coded information might be used by the central auditory system. Recently we have proposed a new kind of neural network, the timing net, that avoids a time-to-place transformation by keeping pitch-related information in the time-domain (Cariani, 2001a). Such nets operate on temporally-coded inputs to produce temporally-coded outputs that bear meaningful information. In this paper, we discuss two areas where temporal codes and neural temporal processing may be relevant to music perception. These involve primitive tonal relations and rhythmic expectancies.

It is possible that many basic tonal relationships are due to the harmonic structure inherent in interspike interval codes. As Licklider (1951) pointed out above, complex tones whose fundamentals are an octave apart (2 : 1) produce many of the same interspike intervals. Other simple frequency ratios, such as the fifth (3 : 2), the fourth (4 : 3), and the major third (5 : 4), also produce intervals in common, the proportion declining as the integers increase. A similarity metric that is based on relative proportions of common intervals thus favors octaves and other simple ratios. Feedforward timing nets extract those intervals that are common across their inputs. In doing so, they carry out neurocomputations for comparing population-wide interspike interval distributions thereby implementing perceptual measures of pattern-similarity. This approach parallels that of Leman and Carreras, except that here the pattern-similarities come directly out of the operation of the neural processing network, without need for prior training or weight adjustments.

In addition to tonal relations, music also plays on the temporal structure of events by building up temporal expectations and violating them in different ways and to different degrees (Epstein, 1995; Jones, 1976; Meyer, 1956). Composers and performers alike use repetition to build up expectations and then use deviations from expected pattern and event timings ("expressive timing") to emphasize both

change and invariance. A very obvious place where strong temporal expectations are created is in the perception of rhythm (Clynes & Walker, 1982; Fraisse, 1978; Jones, 1978; Large, 1994). In the last section of the paper we show how simple recurrent timing nets can build up complex patterns of temporal expectancies on the basis of what has preceded. Such networks may provide basic mechanisms by which auditory images are formed as a stimulus and its associated neural responses unfold through time. They embody simple mechanisms that operate on temporal patterns in their inputs to build up rhythmic expectations which can then be either confirmed or violated. Recurrent time delay networks provide an alternative to temporal processing based on clocks and oscillators.

Our intent in this paper is exploratory rather than systematic, to show some of the potential implications that temporal codes and timing nets might hold for perception of tonal and rhythmic structure in music.

Temporal coding of auditory forms

Temporal codes are neural pulse codes in which relative timings of spikes convey information. In a temporal code, it is temporal patterns between spikes (how neurons fire) that matter rather than spatial patterns of neural activation (which neurons fire most). Temporal coding of sensory information is possible wherever there is some correlation between stimulus waveform and probability of discharge. Such correlation can be produced by receptors that follow some aspect of the stimulus waveform (e.g., phase-locking), such that the stimulus ultimately impresses its time structure on that of neural discharges. Temporal coding is also possible when there are stimulus-dependent intrinsic temporal response patterns (e.g., characteristic response timecourses or impulse responses). In virtually every sensory modality there is some aspect of sensory quality whose perception may plausibly be subserved by temporal codes (Cariani, 1995; Cariani, 2001c; Keidel, 1984; Perkell & Bullock, 1968; Rieke, Warland, de Ruyter van Steveninck, & Bialek, 1997).

Stimulus-driven time structure is especially evident in the auditory system, where a great deal of psychophysical and neurophysiological evidence suggests that such timing information subserves the representation of auditory qualities important for music: pitch, timbre, and rhythm. Of these, a direct temporal code for rhythm is most obvious, since large numbers of neurons at every stage of auditory processing reliably produce waveform-locked discharges in response to each pulse-event.

Population-interval distributions as auditory representations

An account of the neural coding of the pitch of individual musical notes is fundamental to understanding their concurrent and sequential interactions, the vertical and horizontal dimensions of music that contain harmony and melody. To a first approximation, most musical notes are harmonic tone complexes that produce low pitches at their fundamental frequencies. Music theory almost invariably takes the pitch classes of notes as primitive attributes, bypassing the difficult questions of their neural basis. When such foundational issues are addressed within music theory contexts, they are conventionally explained in terms of spectral pattern models, e.g., (Bharucha, 1999; Cohen, Grossberg, & Wyse, 1994; Goldstein, 1973; Parncutt, 1989; Terhardt, 1973).

Spectral pattern theories of pitch assume that precise information about the frequencies of partials is available through prior formation of a "central spectrum" representation. The periodicity of the fundamental, its pitch, is then inferred from harmonic patterns amongst the frequencies of resolved partials. From a neurophysiological perspective, the broadly-tuned nature of neural responses at moderate to high sound pressure levels makes precise spectral pattern analyses based on neural discharge rate profiles across auditory frequency maps highly problematic. In contrast, temporal models of pitch rely on interspike interval information that is precise, largely invariant with respect to level, and found in abundance in early auditory processing.

The two aspects of neural response, cochlear place and time, can be seen in Figure 1. The acoustic stimulus is a synthetic vowel whose fundamental frequency (F0) is 80 Hz. Its waveform, power spectrum, and autocorrelation function are respectively shown in panels A, C, and D. Spike trains of single auditory nerve fibers of anesthetized cats were recorded in response to 100 presentations of the stimulus at a moderate sound pressure level (60 dB SPL) (Cariani & Delgutte, 1996a). The "neurogram" (B) shows the post-stimulus time (PST) histograms of roughly 50 auditory nerve fibers. These histograms plot the probability of occurrence of spikes at different times after the stimulus onset. The most striking feature of the neurogram is the widespread nature of the temporal discharge patterns that are associated with the periodicity of the fundamental. Even fibers whose characteristic frequencies are well above the formant frequency of 640 Hz, around which virtually all of the spectral energy of the stimulus lies, nevertheless convey pitch information. The widespread character of temporal patterns across cochlear frequency territories is a consequence of the broad nature of the low-frequency tails of tuning curves (Kiang et al., 1965). The profile of average driven discharge rates are shown in panel D. The driven rate is the firing rate of a fiber under acoustical stimulation minus its spontaneous discharge rate in quiet. In order to initiate a spectral pattern analysis for estimating the pitch of this vowel, a rate-place representation would have to resolve the individual partials of the stimulus (the harmonics in panel C, which are plotted on the same log-frequency scale as D). In practice, discharge rates of cat auditory nerve fibers provide very poor resolution of the individual harmonics of complex tones, even at very low harmonic numbers. Thus, while there is a coarse tonotopic

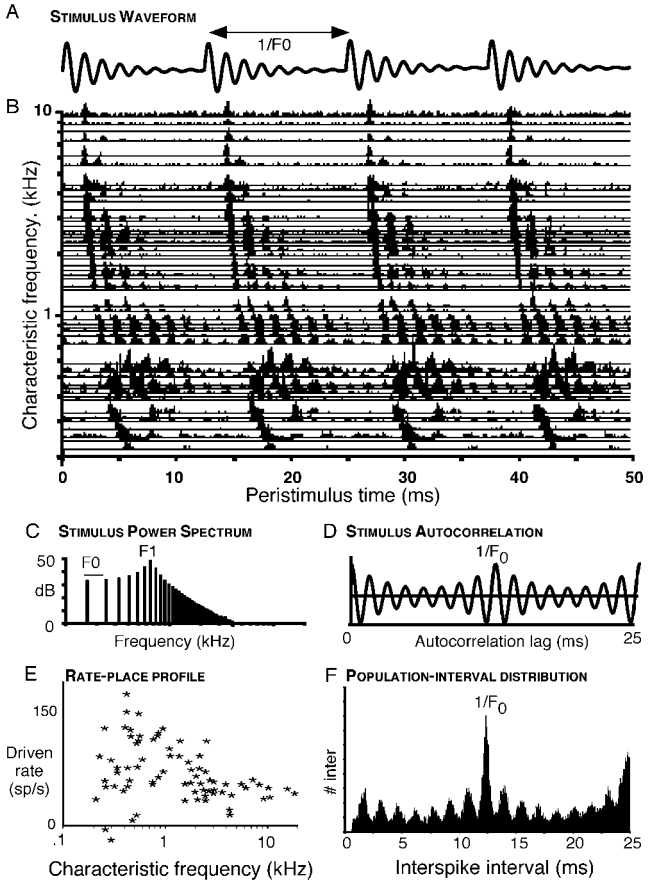

Fig. 1. Temporal coding of musical pitch in the auditory nerve. Auditory nerve responses to a harmonic complex tone with a single formant. (a) Stimulus waveform. A strong, low voice pitch is heard at the fundamental (F0 = 80 Hz, pitch period (double arrow) 1/F0 = 12.5 ms). (b) Peristimulus time histograms of cat auditory nerve fibers (100 presentations at 60 dB SPL). Histogram baselines indicate fiber characteristic frequencies (CF's). (c) Stimulus power spectrum. (d) Stimulus autocorrelation function. (e) Rate-place profile, driven rates as a function of CF. (f) Population-interval distribution formed by summing all-order intervals from all fibers. For further details, see Cariani (1999).

pattern of activity present if one orders the fibers by their characteristic frequencies (cochlear place), this organization is not precise enough to subserve the pitch of complex tones. In contrast, interspike interval information from even a handful of auditory nerve fibers is sufficient to yield reasonably accurate estimates of the fundamental. Pooling interval information from many fibers integrates information from all frequency regions and yields still more precise representations. The population-interval distribution (F) of the ensemble of fibers is formed by pooling all of the interspike intervals from the spike trains produced by the individual fibers. These interspike intervals include time intervals between successive and nonsuccessive spikes, i.e., both "first-order" and "higher-order" intervals are pooled together to form "all-order" interval distributions. Making histograms of all-order intervals is formally equivalent to computing the autocorrelation of a spike train. The population interval histogram (F) shows a very clear peak that corresponds to the fundamental period. For harmonic complexes such as this, the voice pitch that is heard would be matched to a pure tone with the same period, i.e., the pitch is heard at the fundamental frequency. Because of cochlear filtering and phase-locking, the form of the population-interval distribution (F) resembles that of the stimulus autocorrelation function (D) (Cariani, 1999). On the basis of such histograms that contain on the order of 5000 intervals, the fundamental period for such a harmonic tone complex can be reliably estimated, with a standard error of less than 1% (Cariani & Delgutte, 1996a).

Many other detailed correspondences between patterns of human pitch judgment and these global all-order interval statistics of populations of auditory nerve fibers have been found in models, simulations and neurophysiological studies (Cariani, 1999; Cariani & Delgutte, 1996a; Cariani & Delgutte, 1996b; Lyon & Shamma, 1996; Meddis & Hewitt, 1991a; Meddis & Hewitt, 1991b; Meddis & O'Mard, 1997; Slaney & Lyon, 1993). Features of population-interval distributions closely parallel human pitch judgments: the pattern of the most frequent all-order intervals present corresponds to the pitch that is heard, and the fraction of this interval amongst all others corresponds to its strength (salience). Regular patterns of major interval peaks in population-interval distributions encode pitch, and the relative heights of these peaks encode its strength. Many seemingly-complex pitch-related phenomena are readily explained in terms of these population-interval distributions: pitch of the missing fundamental, pitch equivalence (metamery), relative phase and level invariance, nonspectral pitch, pitch shift of inharmonic tones, and the dominance region.

Intervals produced by auditory nerve fibers can be either associated with individual partials or with the complex waveforms that are created by interactions of partials. The first situation dominates at low frequencies, when there is strong phase-locking to the partials (<2 kHz), and for low harmonic numbers, when there is proportionally wider separation between partials. This is the case that is most relevant to musical tones. Here intervals are produced at the partial's period and its multiples, i.e., intervals at periods of its subharmonics. Since all harmonically-related partials produce intervals associated with common subharmonics, at the fundamental and its subharmonics, the most common interspike intervals produced by an ensemble of harmonics will always be those associated with the fundamental (Cariani, 1999; Rose, 1980). Interval distributions produced by harmonic complex tones thus reflect both the overtone series (patterns of partials present in the acoustic waveform) and the undertone series (patterns of longer intervals present in interspike interval distributions). Finding patterns of most frequent intervals in population-interval distributions then is a time-domain analog to Terhardt's frequency-domain strategy of finding common subharmonics (undertones) amongst the partials. Here, the undertone series is directly present in patterns of longer intervals.

In the second situation, when there is weak phase-locking to individual partials (>2 kHz) and harmonic numbers are higher (partials are proportionally closer together), auditory nerve fibers phase lock more strongly to the composite waveform created by interacting partials. For periodic stimuli, this mode of action also produces the most numerous intervals at its repetition period, the fundamental. This was Schouten's "residue" mechanism for the generation of low pitch, where periodicities at the fundamental were thought to be generated by residual modulations left over from incomplete cochlear filtering (Schouten, 1940; Schouten et al., 1962). For a number of reasons, this second situation is considerably less effective at producing intervals related to the fundamental. The dominance region for pitch (de Boer, 1976) and perhaps also the different perceptual characteristics of pitches caused by psychophysically-resolved vs. unresolved harmonics may be explicable in terms of the competition between the two modes of interval production (Cariani, 1999; Cariani & Delgutte, 1996b).

Thus, if pitch corresponds to the most common interval present, whether generated by the first mode of action or the second, then it will always be heard at the fundamental of a harmonic tone complex. Such a representation produces a pitch at the fundamental even if it is "missing" in the frequency-domain description, i.e., there is no spectral energy directly at F0. Because the representation relies on intervals produced by the entire auditory entire array, it also accounts for the inability of low-pass noise to mask the pitch at the fundamental (Licklider, 1954).

Timbre is influenced by spectral energy distribution and by temporal dynamics (e.g., attack, decay). By virtue of phase-locking, both aspects of timbre have neural correlates in the temporal discharge patterns of auditory neurons. Different spectral envelopes produce different interspike interval distributions, since each partial produces intervals according to its relative intensity. Timbres of stationary sounds such as vowels correspond to distributions of short (<5 ms) interspike intervals (Cariani, 1995; Cariani, Delgutte, & Tramo, 1997; Lyon & Shamma, 1996; Palmer, 1992). The pattern of minor peaks in the population-interval distribution of Figure 1 is a reflection of the periodicities of frequency components in the vowel's formant region.

Simulated population-interval distributions

We are interested in how population-interval representations associated with different music notes might be related to each other. A computer simulation of an array of auditory nerve fibers was used to make systematic comparisons between the population-interval distributions that would be produced by different musical sounds. The purpose of the simulation is to replicate the essential temporal features of the auditory nerve response to steady-state signals at moderate to high sound pressure levels in a computationally efficient manner. The MATLAB simulation incorporated bandpass filtering, half-wave rectification, low pass filtering, and rate compression (Fig. 2). Twenty-five frequency channels were simulated with characteristic frequencies (CFs) logarithmically spaced at equal intervals from 100–8000. Each frequency channel contained three classes of auditory nerve fibers, each having its own rate-level function that reflects both spontaneous rate and sound pressure level threshold. Input signals (44.1 kHz sampling rate) were first filtered with a 4th order Butterworth low-pass filter that yields an eight-fold attenuation per octave, and then passed though a 6th order Butterworth high-pass filter that yields three-fold attenuation per octave. Filter and rate-level parameters were chosen that qualitatively replicated the responses of auditory nerve fibers to different frequencies presented at moderate levels (60–80 dB SPL) (Brugge, Anderson, Hind, & Rose, 1969; Kiang et al., 1965; Rose, 1980) and also to the spread of excitation across the array that we observed in our neural data. Consequently, these filters are broader on the low-frequency side than those that are used often used in auditory models that focus on responses at low sound pressure levels, where tuning curves are much narrower. Filtered signals were half-wave rectified and low pass filtered by convolution with a 200 usec square-window. This 200 usec moving average roughly mimics the decline in phase-locking with frequency. Maximal sustained firing rates were then computed for each spontaneous rate class using average root-mean-square magnitudes of the filtered signals. Instantaneous firing rates were computed by modulating maximal sustained rates using the filtered, rectified signal. When the sustained firing rate fell below spontaneous rate in a given channel, uncorrelated, ("spontaneous") activity was generated using a Poisson process whose rate brought the total firing rate up to the baseline, spontaneous rate value. An array of simulated post-stimulus time (PST) histograms was thus generated. Responses of the simulated auditory nerve array (Fig. 2) can be directly compared with the observed neural responses to the same single-formant vowel (Fig. 1). Next, the autocorrelation function of the PST histogram in each channel was computed, and channel autocorrelations were summed together to form the simulated population-interval distribution.

Population-interval distributions and autocorrelation

Simulated population-interval distributions and autocorrelation functions were used to explore pattern-similarities between different notes and chords. Population-interval distribution based on simulated ANFs (Fig. 3, middle column) are compared with those estimated from real neural data (Cariani & Delgutte, 1996a) (left column), and their respective stimulus autocorrelation functions. The positive portions of the autocorrelation functions are shown. For these stimuli the positive portion of the autocorrelation is the same as the autocorrelation of the half-wave rectified waveform.

The four stimuli all produce the same low pitch at 160 Hz: a pure tone (strong pitch, narrow band stimulus), an

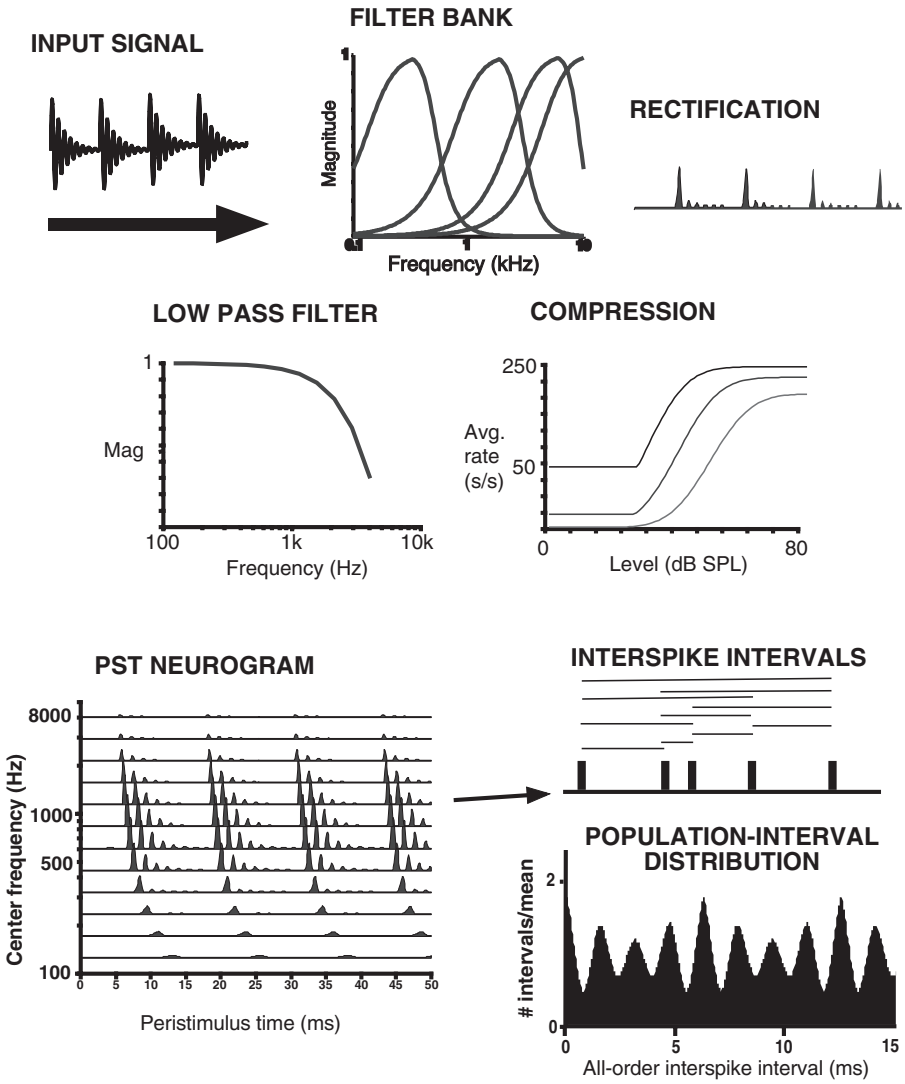

Fig. 2. Auditory nerve array simulation for the estimation of population-interval distributions. An input signal is passed through a bank of bandpass filters, half-wave rectified, low pass filtered, and compressed using three rate-level functions to produce post-stimulus time (PST) histograms for each frequency channel. The autocorrelation of each PST histogram represents its all-order interspike interval histogram. The estimated population-interval distribution is the sum of all channel autocorrelations.

amplitude-modulated (AM) tone (strong pitch, missing fundamental, narrow band), a click train (strong pitch, broadband), and an AM broadband noise (weak pitch). Histogram bins have been normalized by dividing by the histogram mean. The locations and spacings of major peaks in autocorrelation functions and population-interval distributions are virtually the same across the plots, such that these three representations would produce the same pitch estimates. For these stimuli that produce low pitches at 160 Hz, major peaks are located at 6.25 ms and its integer multiples (12.5 ms).

Pitch frequency can be explicitly estimated by finding prominent peaks in population-interval distributions or by examining the repetition pattern of the whole histogram. Earlier estimations involved locating the first major peak in the interval distribution (Cariani & Delgutte, 1996a; Cariani & Delgutte, 1996b; Meddis & Hewitt, 1991a). More recently, we have devised a more satisfying method for estimating pitch that takes into account repeating structure in the whole interval pattern. In this method, all intervals that are part of an interval series are counted, and the pitch is estimated to correspond to the series with the most intervals (highest mean bincount). For example, the sieve corresponding to 200 Hz contains intervals near 5, 10, 15, 20, 25, and 30 ms. This method is more general than peak-picking and is relevant to estimating the relative strengths of multiple pitches that can be produced by multiple interval subpatterns. The relative strength of a given pitch is estimated to be the ratio of the mean bincounts for its sieve to the mean of the whole distribution. The interval sieve is used in this context as an analysis of the all-order interval representations rather than as a hypothetical neural operation. A time-domain theory of pitch multiplicity and of pitch fusion can be built up from such

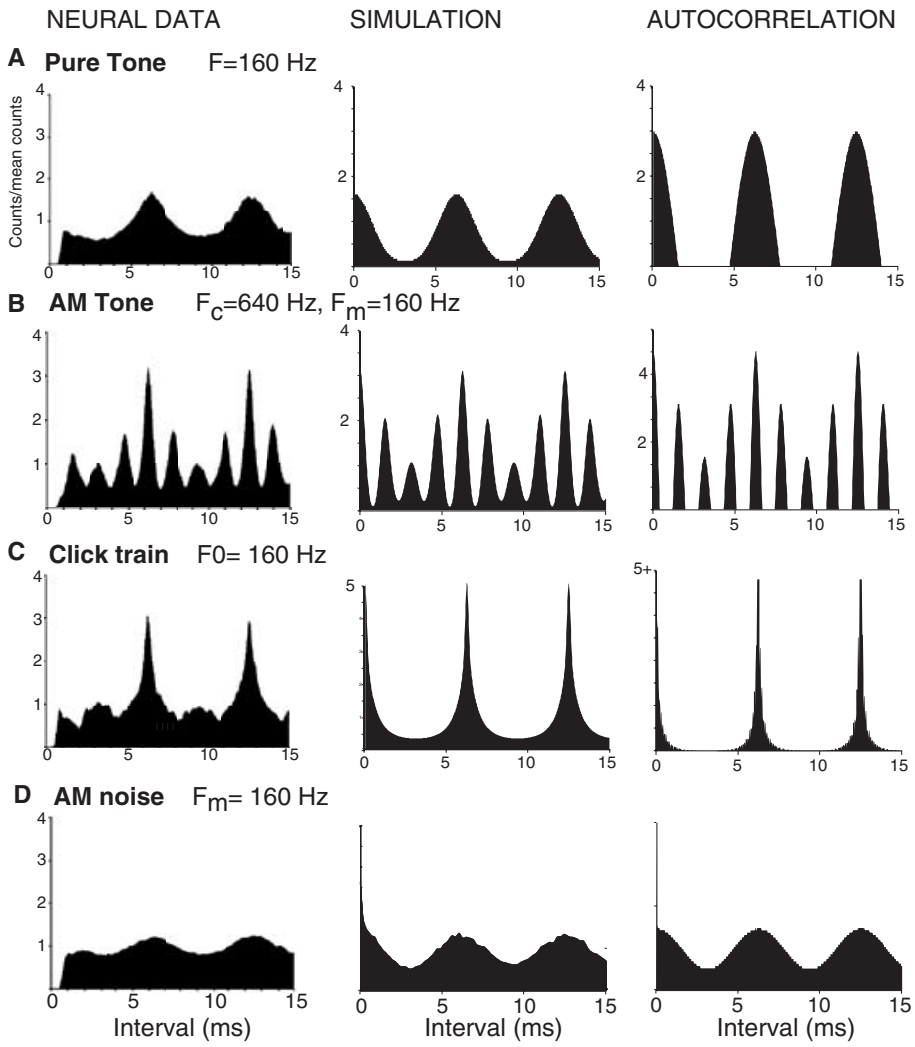

Fig. 3. Comparisons of population-interval distributions and autocorrelation function for six stimuli that produce a low pitch at 160 Hz. Left. Population interval distributions estimated from recorded respones of 50–100 auditory nerve fibers in Dial-anesthetized cats (Cariani & Delgutte, 1996). Middle. Population interval distributions estimated from responses of 75 simulated auditory nerve fibers. Right. Positive portions of stimulus autocorrelation functions.

comparisons of relative pattern strength. Such processes may explain aspects of musical consonance that do not appear to be due to beatings of nearby partials that are associated with roughness (see (DeWitt & Crowder, 1987; Schneider, 1997; Schneider, 2001, in press; Sethares, 1999; Terhardt, 1973; Tramo, Cariani, Delgutte, & Braida, 2001) for discussions).

For our purposes here, we are interested in relations between pitches, e.g., pitch-matching and pitch similarity, rather than absolute estimates of pitch. Our working hypothesis is that the whole interval pattern is itself the neural representation of pitch, and that relative pitch comparisons, which depend on similarity relations, need not depend upon comparisons between prior explicit pitch estimates. These comparisons do not depend on peak-picking or sieve analysis.

In the interval distributions and autocorrelations, those stimuli that produce strong pitches produce high peak-to-mean ratios in population-interval distributions ($p/m > 1.5$), which means that a larger fraction of the intervals that they produce are pitch-related (e.g., at 1/F0 and its multiples). Those stimuli that produce weaker pitches produce lower peak-to-mean ratios ($1.3 < p/m < 1.5$), and those stimuli that fail to produce definite pitches produce ratios close to unity ($p/m < 1.3$).

There are some differences between the three representations. They diverge in 1) the relative heights of their interval peaks, and 2) in the relative numbers of intervals that are not correlated with the stimulus. Relative peak heights differ between the neural systems and their autocorrelation counterparts. This is due to nonlinear processes in real and simulated auditory systems. Were these systems completely linear, population-interval distributions would exactly replicate autocorrelations. Nonlinearities include those generated by cochlear mechanics, threshold and saturation effects in neural rate-level functions, and nonlinear neural membrane

dynamics. In terms of population-interval distributions, nonlinearities have the effect of altering relative heights of interval peaks without changing their positions. The effects that these nonlinearities have on auditory function depends critically on the nature of the neural codes involved. Neural representations for frequency and periodicity analysis that are based on positions of interval peaks rather than numbers of spikes produced are particularly resistant to such nonlinear processes (Cariani et al., 1997). Uncorrelated spikes also produce divergences between the plots. Auditory nerve fibers endogenously produce spikes in the absence of any external stimulus ("spontaneous activity"). In quiet, most fibers have spontaneous firing rates above 20 Hz, with some above 100 Hz. At high sound pressure levels, nearly all spike times are correlated (phase-locked) with the stimulus waveform. In between, there is a mixture of endogenous and stimulus-driven spike generation that produces varying degrees of correlation between spikes and stimulus. Uncorrelated spikes produce flat all-order interval distributions, so that the effect of endogenously-produced spikes is to raise the baseline of the population-interval distribution. One sees the presence of these endogenously produced intervals most clearly by comparing baseline values for stimuli A–C. The neural data shows the highest baselines, the autocorrelation function shows the least, and the simulated cases lie in between. What this shows is that the neural simulation currently captures some of the "internal noise" of the system, but not all of it. As a consequence, the simulation tends to overestimate the fraction of pitch-related intervals produced by the auditory nerve array amongst all other intervals. This fraction is in effect a signal-to-noise ratio for an interval code that qualitatively corresponds to pitch salience (Cariani & Delgutte, 1996a).

The population-interval distribution is a general-purpose auditory representation that generally resembles the autocorrelation function of the stimulus (compare Figure 1D and F). Formally, the autocorrelation function of a stimulus contains the same information as its power spectrum. Thus, to the extent that there is phase-locking to the stimulus, such a representation can subserve the same functions as a frequency map, albeit through very different kinds of neural mechanisms.

Simulated population interval distributions therefore offer rough, but reasonable approximations to interval distributions observed in the auditory nerve. For most pitch estimation purposes involving musical stimuli, the stimulus autocorrelation function would suffice (i.e., bypassing the simulation). The autocorrelation function is thus not a bad first estimate of the form of the population-interval distribution, so long as one is interested in musical pitch (harmonics below 2 kHz) and one's purpose is indifferent to signal-to-noise ratio (i.e., not involving pitch salience, masking, or detectability or competing auditory objects). While there are other special situations that involve higher harmonics and masking effects for which simple autocorrelation models break down (Kaernbach & Demany, 1998), these situations are far removed from those encountered in musical contexts.

Common temporal patterns and pitch similarity

In order to determine whether perceived similarities between musical tones could be based on the similarities of their respective population interval representations, auditory nerve responses to tones with different fundamentals were simulated. Population-interval distributions were compiled from the simulated responses. Pure tones and tone complexes consisting of harmonics 1–6 for fundamentals ranging from 30 to 440 Hz were used as stimuli.

Simulated population interval distributions for a series of fundamental frequencies related by different frequency ratios, including many found in a just-tempered scale are shown in Figure 4. These distributions have all been normalized to their means. Some simple relations are apparent. For both pure and complex tones, the distributions have common major peaks when ratios between fundamentals are near 2 : 1, 3 : 1, 3 : 2, and 4 : 3. These correspond to musical intervals of octaves, twelfths, fifths, and fourths. Distributions for F0 = 100 (1 : 1), 200 (2 : 1), and 300 (3 : 1) share intervals at 10 and 20 ms. Distributions for F0 = 100 and 150 Hz (3 : 2) share intervals at 20 ms, those for 133 and 200 Hz at 15 ms, those for 200 and 300 Hz at 10 and 20 ms. Distributions for F0 = 200 and 167 Hz (4 : 3) share intervals at 20 ms. Fundamental ratios near these values, such as those produced by equal temperament tunings, also produce similar interval overlaps.

Peaks in the population interval distribution narrow as fundamental frequency increases. This is most apparent for the pure tone series, and is ultimately a consequence of the character of auditory nerve phase-locking. The period histogram of an auditory nerve fiber in response to a pure tone resembles the positive portion of the sinuisoidal waveform (Kiang et al., 1965; Rose, 1980). Interspike interval histograms consequently resemble the positive parts of autocorrelation functions. Lower frequency pure tones produce spikes throughout half their cycle, with the consequence that spikes produced by lower frequency components are, in absolute terms, more temporally dispersed than their higher frequency counterparts. This has the effect of making interval peaks produced by lower frequency tones broader.

In these plots and for the analysis of pitch-related pattern similarities, we have weighted intervals according to their duration. Shorter intervals have been weighted more than longer ones. In psychophysical experiments, the lowest periodicities that produce pitches capable of supporting melodic recognition are approximately 30 Hz (Pressnitzer, Patterson, & Krumboltz, 2001). There is other evidence that auditory integration of pitch and timbre takes place within a temporal contiguity window: of 20–30 ms. These include time windows 1) over which pitch-related information is integrated (White & Plack, 1998), 2) over which nonsimultaneous harmonics produce a pitch at their fundamental (10 ms)

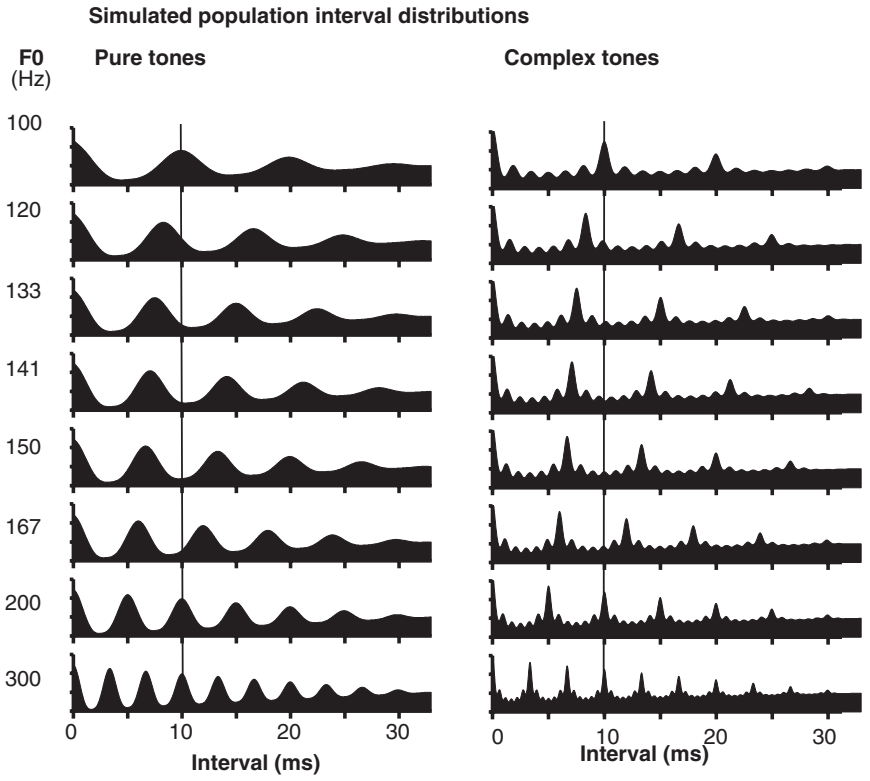

Fig. 4. Similarities between population-interval representations associated with different fundamental frequencies. Simulated population-interval distributions for pure tones (left) and complex tones (right) consisting of harmonics 1–6.

(Hall III & Peters, 1981), 3) over which timbres fuse to produce unified vowels (15–20 ms, Chistovich, 1985; Chistovich & Malinnikova, 1984) or masking of rhythmic patterns (Turgeon, 1999; Turgeon, Bregman, & Ahad, in press), and 5) over which waveform time reversal has no effect on pitch or timbre (30 ms, Patterson, 1994).

To account for the lower limit of pitch, Pressnitzer et al. incorporated a 33 ms window with linearly-decaying weights into their pitch model. The window embodies the assumptions that the pitch analysis mechanism can only analyze intervals up to a given maximum duration (33 ms) and that pitch salience successively declines for progressively lower periodicities (smaller numbers of long intervals). We assume that pitch salience is a function of peak-to-mean ratio in population-interval distributions rather than absolute numbers of intervals, so that a slightly different weighting rule that asymptotes to unity has been used here, $X_w(tau) = 1 + (X(tau) - 1) * (33 - tau)/33$ for all interval durations (taus) 33 ms. This weighting rule reduces the peak to mean ratio of longer intervals. The linear form of the window is provisional, and it may be the case that different periodicities have different temporal integration windows (Wiegrebe, 2001).

Population interval distributions for pure and complex tones are systematically compared in Figure 5. Pearson correlation coefficients (r) between all pairs of simulated population interval distributions associated frequencies from 30–440 Hz are plotted in the upper panel (A). For pure tones (left correlation map) the highest correlations (darkest bands) follow unisons, octaves, and twelfths. For complex tones (right correlation map) there are also additional, fainter bands associated with fifths, fourths, and sixths. Cross sections of the two correlation maps are shown in the bottom panel, where the relative correlation strengths of all frequency ratios can be seen for a few selected notes.

The reason that the population interval distributions show octave similarities lies in the autocorrelation-like nature of these representations (Cariani, 1997, 1999). The autocorrelation of any sinusoidal waveform, irrespective of phase, is a cosine of the same frequency. The unbounded autocorrelation functions of infinitely long sinusoids of different frequencies have zero correlation. However, if waveforms are half-wave rectified and autocorrelation functions are limited by maximum time lags, then these lag-limited autocorrelations of half-wave rectified pure tones will show positive correlations between tones that are octaves apart. Octave similarities between pure tones would then ultimately be a consequence of half-wave rectification of signals by inner hair cells of the cochlea and of the longest interspike intervals that can be analyzed by the central neural mechanisms that subserve pitch perception. The reason that population-interval distributions of complex tones show additional correlation peaks has to do with correlations produced by 1) direct spectral overlap, i.e., partials that are common to the two notes and 2) by octave-relations between sets of partials.

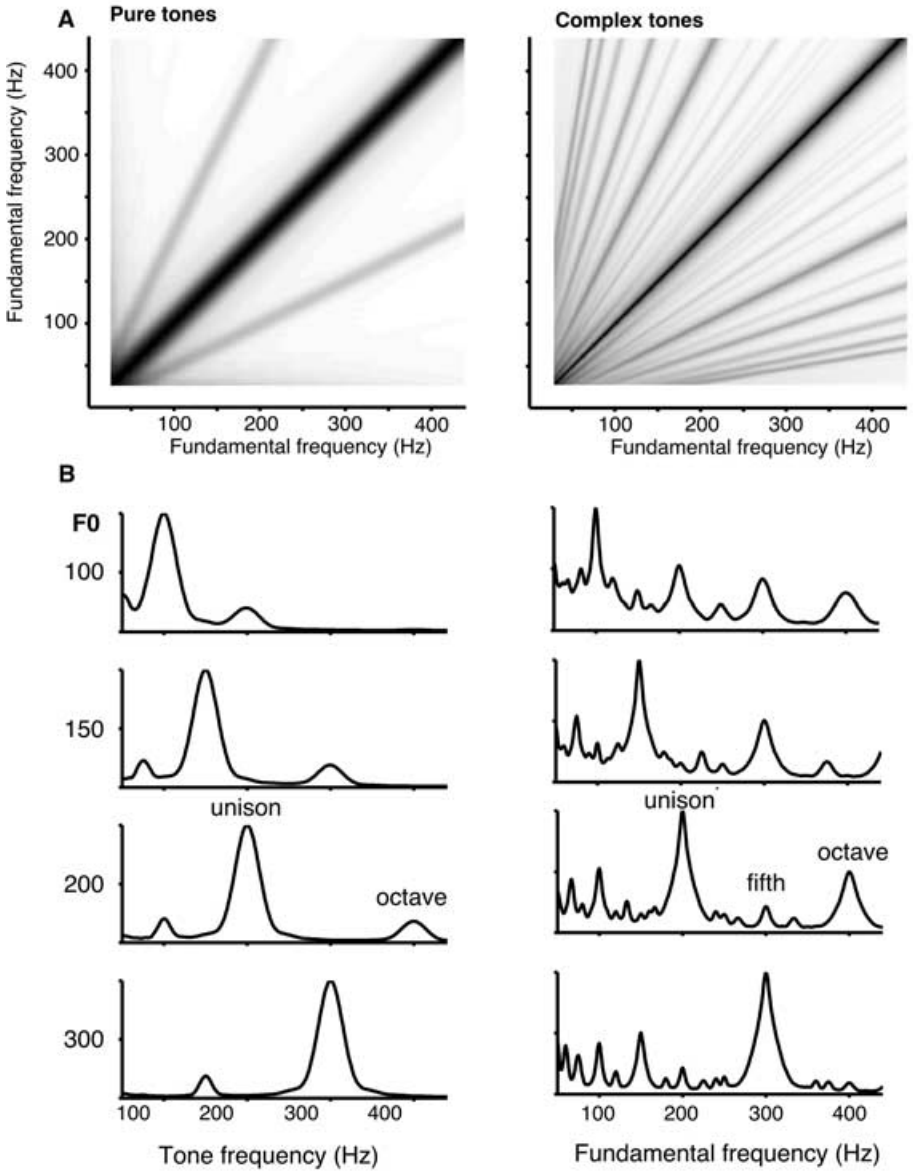

Fig. 5. Tonal structure and pattern similarities between population-interval distributions. Top. Map of correlation coefficients between all pairs of simulated population-interval distributions produced by pure and complex tones with fundamentals ranging from 1–440 Hz. B. Cross-sectional correlation profiles for selected fundamental frequencies. Correlations range from 0–1.

If the auditory system represents pitch through population interval distributions and compares whole distributions to assess their similarity, then by virtue of the properties of these interval-based representations and operations, the system possesses internal harmonic templates that are relativistic in nature. The strongest relations would form structures that would resemble "internal octave templates" (Demany & Semal, 1988; Demany & Semal, 1990; McKinney, 1999; Ohgushi, 1983) in their properties. Octave similarities would then be a direct consequence of neural codes that the auditory system uses rather than of the associative learning of stored harmonic templates or connection weights. The temporal coding hypothesis thus yields a nativist account of basic tonal relations, and provides a means by which complex cognitive schema may be grounded in the microstructure of the neural codes and computations that subserve perception.

For our purposes here we have assumed the correlation comparison as a putative measure of perceptual distance between notes. Here perceptual distance is taken to be inversely related to correlation – those pairs of notes that produce the most interspike intervals in common generate the highest inter-note correlations. According to the interval coding hypothesis, these notes should be the most similar perceptually. Geometrically, zero distances at unisons and the next shortest distances at octaves with distance increasing for successive octaves, translates into a helical structure in which angle corresponds to pitch class (chroma) and distance along the helical axis corresponds to pitch height. Thus the repeating, autocorrelation-like character of all-order interspike

interval distributions produced by periodic sounds can generate both chroma and height dimensions of pitch quality. This ensuing organization of pitch space is consistent with the helical topology that has been inferred from human judgements of pitch similarity (Krumhansl, 1990; Shepard, 1964).

Temporal patterns and note-key relations

One can also assess similarities between interval patterns produced by individual notes and musical chords, and compare these to patterns of similarity judgments by human listeners (Handel, 1989; Krumhansl, 1990; Leman & Carreras, 1997). In a series of studies on tonal context, Krumhansl and colleagues developed a "probe tone" technique for investigating note-note and note-key relationships. In order to minimize possible effects of pitch height, they used notes and note triads made up of octave harmonics in the range from 80–2000 Hz. Notes were constructed in an equally-tempered chromatic scale. Key contexts were established by presenting scales followed by a major or minor triad followed by a probe tone. Experimenters then asked musically experienced listeners to judge how well a particular note "fit with" the previously presented chord. Their averaged, scaled "probe tone ratings" for C major and C minor key profiles are presented in the top left plots of Figure 6 (Krumhansl, 1991, p. 31). Similar judgements are also obtained using other stimuli and key-contexts, so these note-key relationships appear to be general in that they do not depend on particular, familiar key contexts.

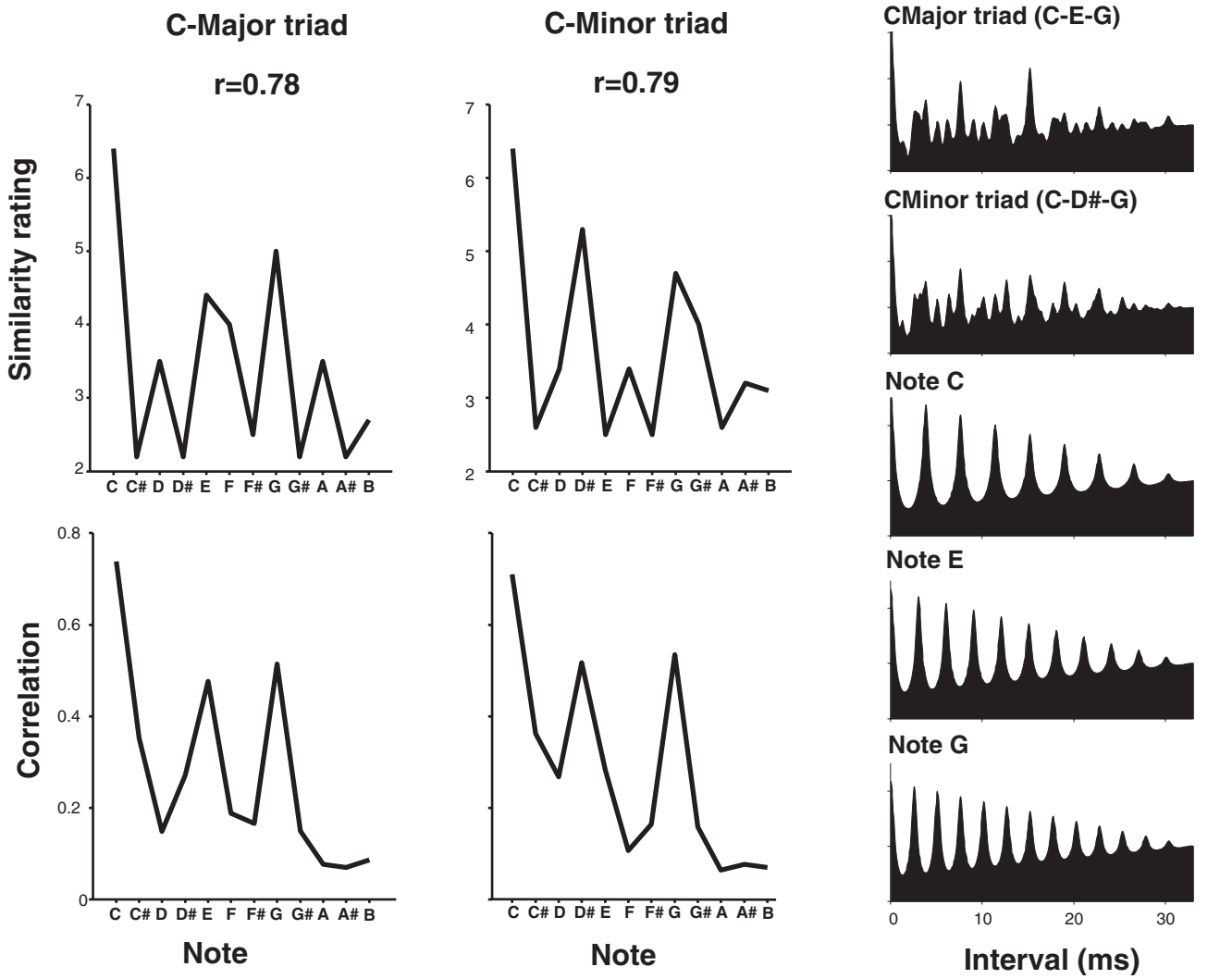

Fig. 6. Comparison of interval-based measures of note-chord similarity with human judgments. Left top. Results of probe tone experiments: human ratings of how well a note fits in with a preceding chord (Krumhansl, 1990). Chords were either C-major (CEG) or C-minor (CD#G) note triads. Notes consisted of harmonics 1–12 taken from an equally-tempered scale. Left bottom. Estimates of tonal similarity based on correlations between simulated population interval distributions. Right. Simulated population interval distributions for the two chords and three individual notes.

As in the corresponding probe-tone studies, notes consisted of harmonics 1–12 of equally-tempered fundamentals, with A set to 440 Hz. Chords consisted of note triads C-E-G (C major) and C-D#-G (C minor). Auditory nerve responses were simulated for the twelve notes and two chords, and their respective population interval distributions were compiled, normalized and weighted as before. The correlation coefficients between all note-chord pairs are shown in the bottom plots on the left. Note-chord similarity profiles were then compared with the probe tone data. Moderately high correlations between the two profiles were observed (r = 0.78 for C-major and r = 0.79 for C-minor). Similar results were also obtained using only octave harmonics and for just-temperament scales. Previously this analysis had been carried out with the unweighted autocorrelations of the notes and chords, with a maximum lag of 15 ms. In this case the correlations were slightly higher (r = 0.94 for C-major and r = 0.84 for C-minor) than for the present, simulated case. Whether population-interval distributions of autocorrelations were used, similarities between these temporal representations of notes and chords paralleled the similarity judgements of human listeners. These results are generally consistent with those obtained by Leman and coworkers (Leman, 2000; Leman & Carreras, 1997), in which chord-chord relations were analyzed using temporal autocorrelation representations. These outcomes are not surprising, considering that population-interval distributions and autocorrelation functions reflect the frequency content of their respective signals and that correlations between them are influenced by both spectral overlap and by octave similarities.

In practice, neurally-based comparison of successively presented chords and notes requires a storage and readout mechanism of some sort. Hypothetically, interval patterns could be stored in a reverberating echoic memory similar in operation to the recurrent timing nets that are discussed further below.

Overtones and undertones in autocorrelation-like representations

The weighted population-interval histograms of the two chords and two individual notes are shown in the panels on the right. The roots of chords, like the low pitches of harmonic complexes, produce patterns of major peaks in autocorrelation-like representations. These kinds of temporal representations seamlessly handle both harmonic and inharmonic patterns. Note that C-major and C-minor have major peaks at 7.6 ms, the period of their common root, C3 (132 Hz). Each chord pattern contains the interval patterns of its constituent notes. Each note pattern in turn is approximately the superposition of the intervals of each of its constituent harmonics. Because the autocorrelation of any periodic pattern contains intervals associated not only with the pattern's repetition, but also those associated with multiple repetition periods, the autocorrelation function contains subharmonics of each frequency component, and by superposition, subharmonics of fundamental frequencies. The same arguments also apply to population-interval distributions (PID's). For example, a pure tone at 1000 Hz produces many all-order interspike intervals at 1 ms and its multiples, such that the interval peaks are located at 1,2,3, . . . ms lags. The note PID's in Figure 6 show peaks at fundamental periods and their subharmonics. In this sense autocorrelation and population-interval representations contain both overtones (harmonics) of musical sounds, because they are present in the acoustics, and their undertones (subharmonics), because they are periodic. A few explanations for the roots of chords based on undertone series have been raised in the past (Makeig, 1982), including Terhardt's algorithm for inferring virtual pitch from the subharmonics of frequency components (Terhardt, 1979). Although schemes based on subharmonics have been dismissed by music theorists (Hindemith, 1945) on grounds that they have are no apparent representation in auditory frequency maps, clearly such a representation can be found in patterns of longer interspike intervals.

Implications for musical tonality

Temporal codes in the auditory system may have wide ranging implications for our understanding of musical tonal relations. If the auditory system utilizes interspike interval codes for the representation of pitches of harmonic complexes and their combinations, then basic harmonic relations are already inherent in auditory neural codes. Basic musical intervals that arise from perceptual similarity – the octave, the fifth, the fourth – are then natural and inevitable consequences of the temporal relations embedded in interspike intervals rather than being the end result of associative conditioning to harmonic stimuli. No ensembles of harmonic templates, be they of harmonics or subharmonics, need be formed through learning. Rather than proceeding from a tabula rasa, learning mechanisms would begin with a basic harmonic "grammar" given by the interval code and elaborate on that. Thus there is a role for the learning of musical conventions peculiar to one's own culture as well as refinement and elaboration of musical perception, but these occur in the context of universally shared faculties for handling basic harmonic relations (Tramo, 2001). Many animals plausibly posses these universal faculties (Gray et al., 2001), since fish, amphibia, reptiles, birds, and mammals hear pitches at the ("missing") fundamentals of tone complexes (Fay, 1988) and have phase-locked neural responses that support interspike interval coding of such periodicities (e.g., Langner, 1983; Simmons & Ferragamo, 1993).

In the last few decades, in the midst of the Chomskian revolution in linguistics and the rise of the digital computer, symbolic, rule-based mechanisms were used to account for much of the tonal and rhythmic structure of music. In this present account, basic cognitive structures can arise from temporal microstructures of auditory perception. Here the perceptual representations are analog and iconic in charac-

ter, mirroring in many ways the acoustic waveform. An interval code is an analog code – although the spike events that delimit the interval are discrete, the time interval itself can take on a continuous range of durations. Even though the representations can vary continuously, their similarity relations partition the space of possible periodicities to form discrete regions that correspond to basic musical intervals (octaves, fifths, fourths). Out of the continuous dynamics of analog representations arise the symbols of rule-based descriptions (Cariani, 2001b). This is perhaps a means by which Kohler's hypothesis of perceptual isomorphism (Boring, 1942; Leman & Carreras, 1997) can accommodate both continuous qualities, such as pitch, timbre, tempo, as well as discrete categories, such as discernable musical intervals and classes of rhythmic patterns.

A major question for temporal codes involves the means by which the auditory system would make use of such information. In the second half of the paper we present a new kind of neural network, the timing net, that operates on temporally-coded inputs. We will show how feedforward timing nets can implement comparisons between population interval distributions, and how recurrent timing networks can build up rhythmic patterns that recur in their inputs.

Neural timing nets

Thus far we have discussed the implications of neural population-based interspike interval codes for musical pitch relations. However, a signal has meaning only by virtue of how it is interpreted by a receiver. In order to bear meaningful informational distinctions, putative neural representations must be interpretable by biologically-embodied neural architectures. Each possible neural code is intimately linked with the neural processing architectures that can interpret it, and each architecture in turn makes assumptions about the nature of the neural signals that it processes. Conceptions of neural networks inevitably embody deep general assumptions about neural codes and vice versa.

Rationale for development

By far the dominant assumption in both neuroscience and music theory is that the auditory system consists of an array of band-pass filters in the cochlea that produce spatial activation patterns in auditory frequency maps that are subsequently analyzed by connectionist networks (Bharucha, 1991; Bharucha, 1999; Cohen et al., 1994). While it is possible to arrange inter-element connectivities in a manner that permits the pitches of complex tones and their equivalence classes to be computed, in order for these networks to attain discriminative precisions on par with those of humans and animals, their inputs must be highly frequency selective and robust. With a few exceptions, the narrow tunings that are required are generally at odds with those that are seen in the auditory pathway, where neural response areas typically broaden greatly at moderate to high sound pressure levels. Many modelers simply sidestep the issue by using very narrow frequency tunings that are derived from human psychophysical experiments, but this assumes away the mechanisms by which cochlear and neural responses produce fine discriminations in the first place. In the midst of frequency-domain operations on "central spectra" derived from psychophysically-derived auditory filters, it can easily be forgotten that the central spectra themselves may be based on interspike interval information rather than rate-place profiles (Goldstein & Srulovicz, 1977; Moore, 1997a).

We do not at present have an adequate account of how the auditory system actually utilizes such interval information to discriminate pitches produced by pure and complex tone. Arguably, the best neurocomputational models that address this problem are temporal autocorrelation networks in the tradition of Jeffress and Licklider, "stereausis" models (Lyon & Shamma, 1996), and modulation-analysis networks (Langner, 1992). A notable recent proposal that uses temporal patterns and cochlear phase delays to tune a coincidence network is that of (Shamma & Sutton, 2000). All of these networks carry out a time-to-place transformation in which information latent in interspike intervals and neural synchronies is converted into an across-neuron activation pattern that can be subsequently analyzed by central connectionist networks. To this end, temporal autocorrelation models use tapped neural delay lines, stereausis models use cochlear delays, and modulation-analysis models use periodicity tuning properties based on neural inputs and intrinsic recovery dynamics.

There are difficulties, however, with each of these schemes. Thus far, auditory neurophysiology has yet to discover any neural populations whose members have (comb filter) tuning characteristics that would be associated with autocorrelating time-to-place architectures. While some central auditory neurons are sensitive to particular pure tone combinations, concurrently and sequentially (Weinberg, 1999), tuning curves and response patterns generally do not betray highly precise harmonic structure commensurate with the precision of the pitch percept itself. Perhaps the most plausible of these schemes given our current state of neurophysiological knowledge is the modulation-analysis hypothesis (Langner, 1992; Schreiner & Langner, 1988). Neurons that are sensitive to particular periodicity ranges are found in abundance at many levels of auditory processing, but their selectivity is coarse and declines at high stimulus levels. A more fundamental, theoretical problem with this hypothesis is that the structure of pitch judgements for harmonic and inharmonic stimuli with low harmonics follows an autocorrelation-like pattern (de Boer, 1956; de Boer, 1976), "de Boer's rule", rather than the pattern that would be produced by a modulation-analysis (Slaney, 1998).

One does find these requisite representational properties in the time domain, in all-order interspike interval distributions. This information is precise, robust, reliable, and appears in great abundance at all auditory stations up to the midbrain and possibly higher. Temporal response patterns observed

Table 1. General types of neural networks.

Type of network	Inputs	Outputs
Connectionist	Channel-coded	Channel-coded
Time delay	Temporally-coded	Channel-coded
Timing net	Temporally-coded	Temporally-coded

from the auditory nerve to the midbrain do follow de Boer's rule (Cariani, 1995; Cariani & Delgutte, 1996b; Greenberg, 1980). The real problem then is to explain the mechanisms by which timing information is utilized in subsequent central auditory processing. Any time-to-place transformation is likely to lose representational precision; pitch estimates based on rate-based tunings are inevitably one or two orders of magnitude coarser than those based on spike timing. For these reasons, alternative kinds of neural networks have been explored that obviate the need for time-to-place conversions by operating completely in the time domain.

Types of neural networks

If one divides neural pulse codes into channel-based codes and temporal codes, then neural networks naturally fall into three classes: 1) those that operate strictly on channel-activation patterns, 2) those that interconvert temporal and channel patterns, and 3) those that operate strictly on temporal spike patterns. Neural architectures of these types can be called, repectively, connectionist networks, time-delay neural networks, and neural timing nets.

Traditionally, neural networks have been conceptualized in terms of spatialized activation patterns and scalar signals. Conventional connectionist nets generally assume synchronous inputs whose time structure is generally irrelevant to the encoding of information. Whatever relevant temporal structure exists is converted to spatial activation patterns by means of temporal pattern detectors.

Time-delay architectures were among some of the earliest neural networks intended to account for the mechanics of perception (Jeffress, 1948; Licklider, 1951). Time-delay neural networks consist of arrays of tapped delay lines and coincidence coinicidence counters which convert fine temporal structure in their inputs spatialized activation patterns in their outputs. The strategic assumption is that temporal patterns are first converted into channel activation patterns, and then subsequently analyzed via connectionist central processors.

Recently we have proposed a third kind of neural network, called a timing net (Cariani, 2001a,d). Timing nets are neural networks that use time-structured inputs to produce meaningful time-structured outputs. Although they share many common structural elements with time-delay neural nets (coincidence detectors, delay lines), timing nets are functionally distinct from time-delay networks in that the goal of the network is to produce a temporal pattern as its output rather than a spatial pattern of element-activations. Time-delay nets use coincidence detectors that are subsequently coupled with an integration or counting mechanism to effect "coincidence counters" that eliminate the temporal information present in the coincidences themselves. Instead, timing nets produce these temporal patterns of coincidences that then can be analyzed by other timing nets.

As with other kinds of networks, timing networks can further be divided into feedforward and recurrent networks on the basis of whether the network contains internal loops. Feedforward timing nets (Fig. 7a) act as temporal pattern

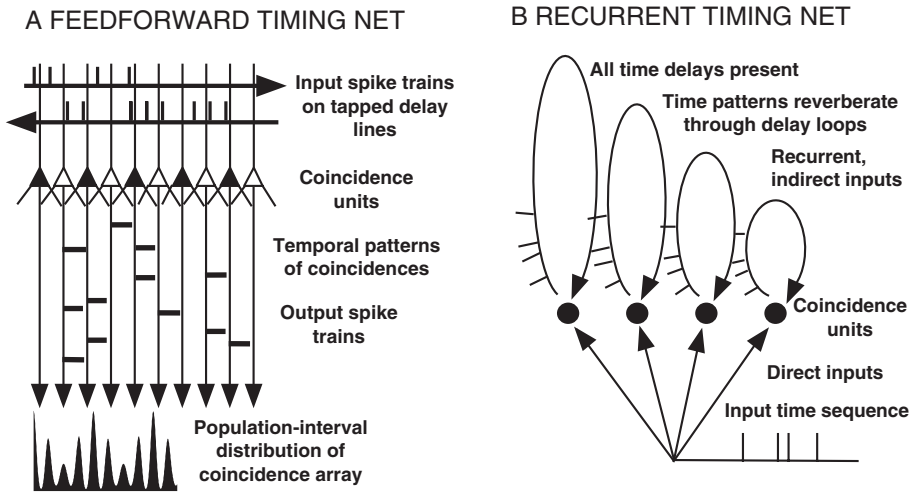

Fig. 7. Neural timing nets. (a) A simple feedforward timing net consisting of two tapped delay lines and a linear array of coincidence detectors. Outputs of coincidence detectors contain only temporal patterns that are common to the two inputs. The population interspike interval distribution of the outputs of the net reflects a comparison between the interval distributions of the two inputs. (b) A simple recurrent net consisting of an array of coincidence detectors fed by direct inputs and by delay loops of different time durations. These networks compare incoming temporal patterns with previous ones to build up temporal expectations.

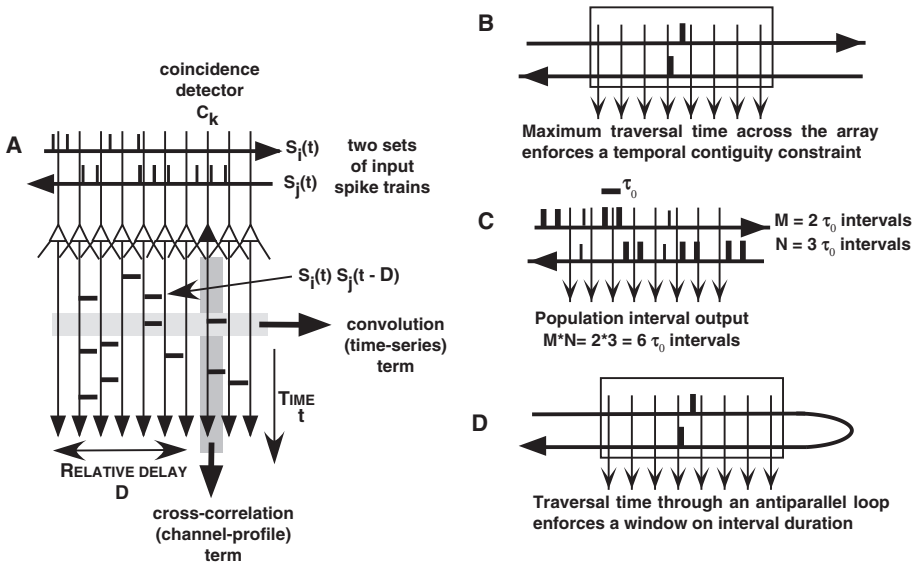

Fig. 8. A simple feedforward timing net. (a) General schematic of a coincidence array traversed by tapped delay lines. Summation over time in each output channel yields the cross-correlation function, while summation over output channels for each time yields the convolution of the two inputs. (b) The population-interval (summary autocorrelation) of the entire output ensemble computes the product of the autocorrelations of the two input channels. (c) The conduction time across the array determines the temporal contiguity window between its inputs. (d) A delay line looped back upon itself produces intervals that are limited by the traversal time of the loop.

sieves to extract common periodicities in their inputs, and thus are relevant to perceptual comparisons of pitch, timbre, and rhythm. Recurrent timing nets (Fig. 7b) build up temporal patterns that recur in their inputs to form temporal expectations of what is to follow. We discuss how recurrent timing networks may be applied to the formation of rhythmic expectations in the last section of the paper.

Timing networks were directly inspired by several temporal processing architectures. Feedforward timing networks are related to the Jeffress temporal cross-correlation architecture for binaural localization (Jeffress, 1948), Licklider's temporal auto-correlation architecture for pitch (Licklider, 1951; Licklider, 1959), the combination auto- and cross-correlation architectures of Licklider and Cherry (Cherry, 1961; Licklider, 1959), Braitenberg's cerebellar timing model (Braitenberg, 1961), and the temporal correlation memory strategies suggested by Longuet-Higgins (Longuet-Higgins, 1987; Longuet-Higgins, 1989). Although feedforward networks of all kinds have been studied in greater depth because of their formal tractability, in view of the ubiquity of reciprocal connectivities between neurons and neural populations, theoretical neuroscientists have always looked to recurrent networks as more realistic brain models. Thus, adaptive resonance circuits, re-entrant loops, and thalamocortical resonances are prominent in the current thinking about large scale neural integration. While conceptions incorporate notions of reverberating circuits (Hebb, 1949), these schemes use spatial patterns (Grossberg, 1988) and sequences of neural activations (e.g., McCulloch, 1969), rather than temporal codes. Nonetheless there have been a few proposals for temporal processing using neural delay loops (Thatcher & John, 1977). In the auditory system, Patterson's strobed temporal integration model (Patterson, Allerhand, & Giguere, 1995) functions in a manner similar to a recurrent timing network in that it retains previous temporal patterns that are then cross-correlated with incoming ones to build up stable auditory images. The timing networks that we describe here are not adaptive networks that adjust inter-element connectivities (synaptic weights) or delays (conduction times), but such adaptive mechanisms have been proposed in the past (MacKay, 1962), and are important to any general hypothesis concerned with neurocomputational substrates of learning and memory.

Feed-forward timing networks

The simplest kind of timing net is a feed-forward network. In such a network, two pulse train time-series signals ($S_i(t)$ and $S_j(t)$) are fed into a coincidence array via two tapped delay lines (Fig. 8a). Whenever a coincidence element receives a nearly simultaneous pulse from each set of lines, it produces a pulse in its output. Each channel, by virtue of its position in the array relative to the two input delay lines computes the pulse correlation at a specific relative inter-element delay (D). A pulse appearing in the output of a given element C_k therefore reflects the conjunction of two pulse events whose times of occurrence are separated in time by its characteristic relative delay D_k. For pulse train signals, the output of a particular detector $C_k(t)$ is equal to the product of the two binary signals (0: no pulse, 1: pulse) at the detector's characteristic delay, $S_i(t)S_j(t - D_k)$.

Basic computational properties

Two kinds of functions can be computed if the outputs of the array are summed together in channel or in time. Integrating the activity for each channel over time (vertical shaded area) computes the cross-correlation function of the two inputs. Adding this coincidence counting operation (and dividing by the integration time to compute a running coincidence rate) makes the network functionally equivalent to the Jeffress architecture for binaural cross-correlation. Here the channel activation profile of the coincidence elements (i.e., a rate- "place" pattern) represents the cross-correlation. Integrating delay channels for each time step (the horizontal shaded area) computes the convolution of the two signals (Longuet-Higgins, 1989). Thus the population-wide peristimulus time (PST) histogram of the ensemble of coincidence elements reflects the convolution of the two inputs. These operations, however, do not exhaust all the functional possibilities.

The time structure of the coincidences themselves bear a great deal of useful information. In essence, feed-forward timing nets act as temporal-pattern sieves, passing through to the individual channel outputs those temporal patterns that the two input signals have in common. For a pulse to appear in somewhere in the array's output, a pulse must have been present in each input at roughly the same time, i.e., within the temporal contiguity constraint imposed by the travel time across the array. Two pulses arriving outside this contiguity window do not cross in the array. For a particular interspike interval to appear in the output of a coincidence element in the array, that interval must have been present in the two inputs. The same argument holds for higher-order patterns, such as spike triplets and longer sequences: if one observes a higher order pattern in at least one of the output channels, then the pattern must have been present in both inputs.

One desires a means of representing this information that is latent in the array's output. We will explore the behavior of the interval distribution produced by the ensemble of coincidence detectors, i.e., its population-interval distribution. The autocorrelation function of a spike train is formally equivalent to its all-order interspike interval histogram. The autocorrelation of the output of a particular coincidence detector C_k is $A_k(tau) = \Sigma[S_i(t)S_j(t - D_k)][S_i(t)S_j(t - D_k) - tau)]$. This is the product of the output of the detector and delayed copy of itself summed over time (t) for each of many delays (tau: interval duration). Summing together the temporal autocorrelations of the output from each of the elements in the coincidence array produces the summary autocorrelation of the entire array, i.e., $SAC(tau) = \Sigma A_k$. In neural terms this is the population-interval distribution of the coincidence array, i.e., the global all-order interval statistics of the whole ensemble of coincidence elements.

The traversal time across the array determines which parts of the signals interact with each other (Fig. 8b). All intervals from each set of inputs that arrive within the temporal contiguity window cross their counterparts in the other set, such that if one input has M such intervals of duration tau, and the other has N such intervals, $M * N$ tau-length intervals will appear in the outputs (Fig. 8c). Within the temporal contiguity constraints, the coincidence array therefore performs a multiplication of the autocorrelations of its inputs. Thus, for any pair of signals, if we want to know the all-order population-interval distribution (summary autocorrelation) that is produced by passing them through such an array, we can multiply their all-order interval distributions (signal autocorrelations). If an input line is looped back upon itself in antiparallel fashion to form a recurrent loop (Fig. 8c), then the maximum autocorrelation lag that will be computed by the loop is determined by the maximal traversal time of the overlapped segments.

Extraction of common pitch and timbre

The multiplication of autocorrelations has important functional consequences for subserving pitch and timbre comparisons. Feedforward timing nets implement a comparison operation that is related to the correlation-based metric that was used to explore tonal relations (Figs 5 and 6). Coincidence arrays extract all periodicities that are common to their inputs, even if their inputs have no harmonics in common. This is useful for the extraction of common pitches irrespective of differences in timbre (e.g., two different musical instruments playing the same note), and extraction of common timbres irrespective of pitch (the same instrument playing different notes). Here we focus on those aspects of timbre that are associated with the spectral composition of stationary sounds, as opposed to those aspects that have dynamic origins. On longer time scales, but using similar temporal computational strategies, different rhythmic patterns can be compared to detect common underlying meters and subpatterns.

The results of such comparison operations are shown in Figure 9. Four electronically synthesized waveforms differing in the pitches and timbres they produce were recorded using a Yamaha PSR76 synthesizer with different voice settings. Two notes, C3 and D3, and three voices, "Pipe organ" (A), "alto sax" (B) and "sustained piano" (C) were chosen and waveforms were taken from the stationary, sustained portion of the sounds. Their combinations cover different commonalities of pitch (note) and timbre (instrument): AB, common timbre, different pitch; AC, different timbre, same pitch; AD, different timbre, different pitch. Waveforms, power spectra, and autocorrelation functions are shown for the four sounds. Simulated population interval distributions were computed for each of the four waveforms, and each distribution was normalized relative to its mean.

Patterns of major peaks associated with note fundamentals (pitch), and patterns of short-interval minor peaks associated with the instrument voice settings (timbre) are readily seen in the autocorrelations and population interval distributions. The equally-tempered note C3 has a funda-

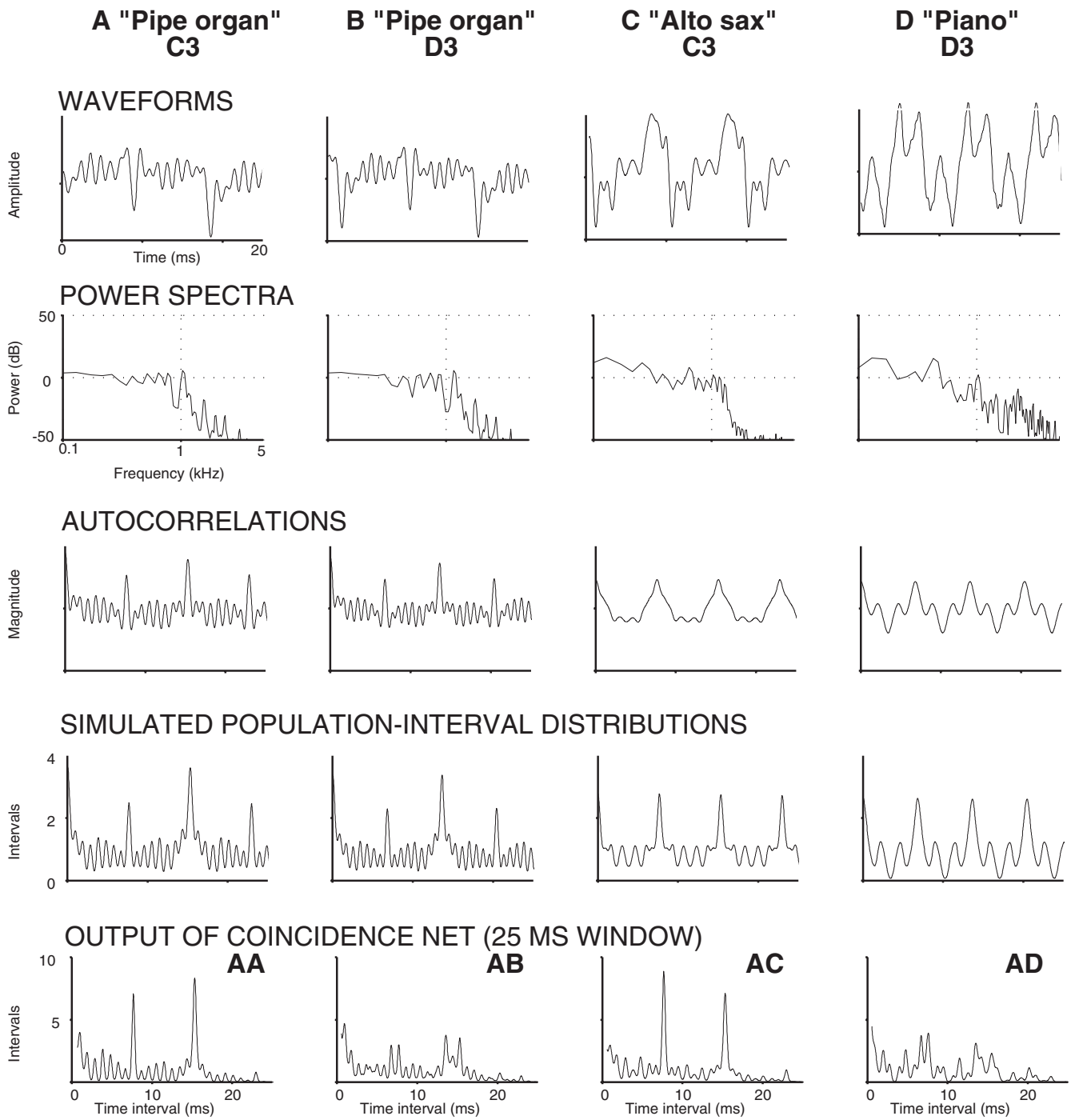

Fig. 9. Pitch matching irrespective of timbral difference. Waveforms, power spectra, autocorrelations, and simulated population-interval distributions are shown for four synthesized waveforms (Yamaha PSR-76). Complex tones A and B have similar spectral envelopes and produce the same timbre. Tones A and C have a common fundamental, and evoke the same musical pitch (C), but have different spectral envelopes and have different timbres. Tones A and D have different fundamentals and spectral envelopes. Windowed products of pairs of population-interval distributions (Tone A paired with A–D) are shown in the bottom row.

mental at 131 Hz or 7.6 ms, while D3 has a fundamental at 147 Hz or 6.8 ms. The correlation coefficients between the four population interval distributions are given in Table 2. Correlation comparisons between population interval distributions heavily weight commonalities of fundamental frequency (pitch) over those of spectral shape (timbre). Correlations between the patterns in the different interval ranges associated with timbre (0–2 ms) and pitch (2–10 ms) of these stimuli are presented in Table 3. The "pipe organ" and "alto sax" voices have very similar short-interval patterns, but these differ substantially from that of the "piano" voice.

Table 2. Correlations between population interval distributions for waveforms A–D.

Note	Frequency (pitch)	Period (pitch)	Voice (timbre)	r	A	B	C	D
C3	131 Hz	7.6 ms	"pipe organ"	A	1			
D3	147 Hz	6.8 ms	"pipe organ"	B	0.10	1		
C3	131 Hz	7.6 ms	"alto sax"	C	0.72	0.06	1	
D3	147 Hz	6.8 ms	"piano"	D	0.08	0.63	0.05	1

Table 3. Correlations for 0–2 ms (bottom) and 2–20 ms intervals (top).

	A	B	C	D
A	1	–0.07	0.69	–0.04
B	0.94	1	–0.08	0.61
C	0.96	0.96	1	–0.03
D	0.78	0.76	0.73	1

The bottom row of plots shows the products of the simulated population interval distributions after a 25 ms tapering window was applied. The product of a population interval distribution with itself is its square (AA), which retains the patterns of major and minor peaks associated with pitch and timbre. The product of distributions associated with common timbre, but different pitch (AB) shows no prominent major peak, but replicates the short interval pattern (0–3 ms) that is common to the two input signals. The product of distributions associated with common pitch, but different timbre (AC) shows prominent major interval peaks at the common pitch period of 7.6 ms. Finally, the product associated with different pitches and timbres (AD) produces no prominent pitch-related peaks and the timbre-related pattern of short intervals resembles that of neither input. A similar analysis has been carried out with four different synthetic vowels that pair one of two fundamentals (voice pitches) with one of two formant configurations (vowel quality) (Cariani, 2001a,d).

The feedforward timing network thus produces an interval distribution in its output that corresponds to the common pitch of the two inputs, and it does this without ever having to explicitly compute the fundamental frequency. This kind of comparison operation produces relative pitch judgements rather than the absolute ones that would be generated from an explicit pitch estimate. The relativistic nature of this periodicity-matching process parallels the relativity of pitch perception. In order to match pitches using this scheme, one adjusts the fundamental frequencies (as in tuning an instrument) so as to maximize the relative numbers of intervals produced by the coincidence array. Here the relative numbers of intervals produced by the stimulus combinations were AA (20), AB (16), AC (18), and AD (15), producing a similarity ordering of 1) common pitch and timbre, 2) common pitch, slightly different timbre, 3) different pitch, slightly different timbre, and 4) different pitch and timbre. This simple strategy of pitch matching by maximizing the output of the whole array works as long as sound pressure levels (and consequently input spike rates) are kept constant, but would be expected to fail if matching were carried out with roving levels. Maximizing the peak to background ratio of intervals in the output that are associated with the pitch periodicity to be matched achieves a normalization operation that makes the computation more like a correlation comparison (as in Figs 5 and 6). Further development of the computation of similarity in timing nets should include incorporation of inhibitory elements that impose penalties for anticoincidences and perform cancellation-like operations (de Cheveigné, 1998; Seneff, 1985; Seneff, 1988).

While the multiplication of autocorrelations is related to the product of power spectra, there are some notable functional differences between the two. A timing network extracts intervals common to two fundamentals even in the case where the two inputs do not have any harmonics in common. For example two amplitude-modulated (AM) tones with the same fundamental but different carrier frequencies produce the same low pitch at their common "missing" fundamental. Intervals related to this periodicity predominate in the output of a timing net (Cariani, 2001a,d). Recognition of this similarity cannot involve spectral overlap, since there is none. Thus pitch matching using frequency representations cannot be accomplished by simple assessment of spectral overlap, and necessarily requires a prior harmonic analysis of component frequencies that makes an explicit estimation of pitch. While the representation of the fundamental is simple and prominent in interspike interval-based representations, in contrast, in spectral pattern representations it is implicit and must be extracted by fairly elaborate means. In addition, the same interval-based representations and neurocomputations can subserve both pitch and timbre comparisons using the same timing net operations, whereas spectral pattern strategies require different types of analyses (spectral overlap and harmonicity).

Population interval distributions and timing nets exhibit a number of properties that embody Gestaltist principles. One aspect of autocorrelation-like representations is that they exhibit properties related to both parts and wholes. The autocorrelation patterns of the partials are present in the whole

pattern, but the whole has properties, a pattern of major and minor peaks, that reflect combinations of partials. Properties of either whole or part can be analyzed at will. Rather than being concatenations of discrete local features, i.e., perceptual atoms, population interval representations are based on interspike intervals, which themselves are relations between events. These kinds of correlational, relational representations constitute general alternatives to perceptual processing by means of local features and decision trees. Like the intervals themselves, the computations realized by timing nets are analog in character and produce output patterns that can have both continuous and discrete qualities.

Beyond this, there are a host of more general neurocomputational implications that timing nets hold for the nature of neural representational and computational systems. Their ability to extract temporal patterns that co-occur or recur in their inputs, even if these are embedded in other spikes, permit different kinds of information to be carried along the same neural transmission lines and separated out. Timing nets are the only kind of neural net to our knowledge that are indifferent to which particular neuron produces which output response. The operation of the temporal sieve does not depend on specific connectivities between particular neurons, as long as the ensemble encompasses a rich set of delays between the signals. Statistical information can consequently be analyzed by neural ensembles and shipped en masse from one region to another. These properties ultimately liberate neural representations from travel over dedicated lines and processing via connections whose relative weightings must be highly regulated. They provide neurocomputational alternatives to "switchboard-based" modes of organization that require specific connectivities (see John, 1972; Orbach, 1998; Thatcher & John, 1977 for critiques and alternatives).

Recurrent timing nets

Time is central to music in two ways – in the form of (synchronic) temporal patterns and as (diachronic) temporal successions of these patterns. Similarly, time comes into music perception in two ways, as the basis of stable musical forms and qualities (pitch, timbre, rhythmic pattern), and as the dynamic evolution of representations and their changes. As we have seen, relations between musical objects such as notes may be mediated by the underlying temporal microstructure of their neural representations. Sounds also unfold over time, and in parallel with their successions of change are perceptual and cognitive representations that similarly evolve in time. Repetition of pattern plays an important role in music, both as a means of building up invariant object-patterns (be they melodies, harmonies, or meters) and as a means of setting up temporal expectations for future events. Feedforward timing nets are relevant to comparison and analysis of temporal pattern, while recurrent timing nets address the history-dependent evolution of representations and expectations.

Basic properties of recurrent timing nets

Recurrent timing nets consist of coincidence arrays with delay lines that form loops (Figure 7b). Recurrent delay lines provide a primitive form of memory in which a time series signal is presented back, at a later time, to the elements that generated it. From the perspective of temporal processing, a conduction delay loop retains the temporal structure of patterns presented to it, from the timing of a single pulse to complex temporal sequences of pulses. This means that recurrent conduction loops have some functional properties that differ from mechanisms, such as clocks, oscillators and simple periodicity detectors, that use elements that do not retain the full temporal pattern.

Recurrent transmission paths can be constructed in a number of ways. Recurrent loops can be monosynaptic or polysynaptic. Monosynaptic delay loops are based on recurrent collaterals within single elements. Polysynaptic loops are cyclic transmission paths that pass through multiple elements of a network. The brain is rich in cyclic polysynaptic paths because of local interneurons, ascending and descending fiber systems in subcortical pathways and reciprocal, "reentrant" connections between cortical areas (McCulloch, 1947). The impressive array of transmission circuits in the hippocampus provides a rich set of recurrent connections and delays that potentially support autoassociative memories of both channel-coded and temporally-coded sorts. Myelinated pathways provide fast conduction and short delays, while unmyelinated fibers provide much longer ranges of delays. Delay loops can be fixed, prewired, or can arise dynamically, from activity-dependent synaptic facilitation processes. Here we explore the behavior and functionalities of the simplest arrays of fixed, monosynaptic delay loops and coincidence detectors, in order to make a very preliminary survey of their potential usefulness in understanding music perception.

Many different delay loops permit an input signal to be compared with itself at different previous times. If delay loops are coupled to coincidence detectors, then the detectors register correlations between present and past values of the signal, such that the ensemble in effect computes a running autocorrelation. If the outputs of coincidence detectors are fed back into the loop, then an iterated, running autocorrelation is computed that reflects the recent history of both signal and system.

A simple recurrent timing net with these properties is shown in Figure 10A. The behavior of the net was initially explored using binary pulse trains of 0's and 1's. In the absence of a indirect signal coming from within the loop, the loops convey the incoming direct signal. The incoming direct signal is multiplied by the circulating signal and a facilitation factor (B = 1.05). Thus, whenever there is a coincidence of pulses (those arriving from outside with those arriving through the loop), the magnitude of the pulse entering the loop is increased by 5%. Such a network creates a temporal expectation in the timing of future

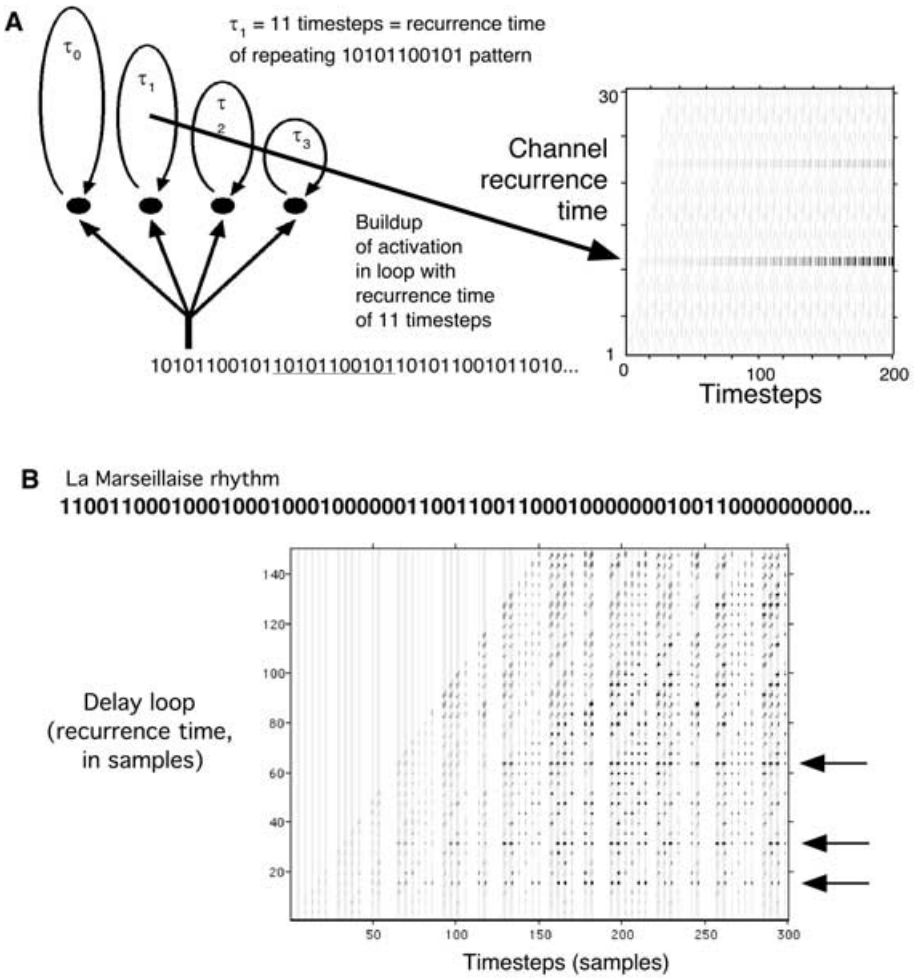

Fig. 10. Behavior of recurrent timing nets. (a) Behavior of a simple recurrent timing net for periodic pulse train patterns. The delay loop whose recurrence time equals the period of the pattern builds up that pattern. (b) Response of a recurrent timing net to the beat pattern of *La Marseillaise*. Arrows indicate periodic subpatterns at 16, 32, and 64 timesteps that are built up by the network.

incoming pulses, and builds up that expectation when it is fulfilled.

Thus, taking sequences of pulses into account, if the incoming signal pattern is similar to the previous pattern, then this pattern is built up within the loop. If the incoming signal pattern is different from the previous pattern, then a new pattern is nucleated within the loop and the build up process begins anew. In effect, an incoming pattern becomes its own matched filter pattern-detector. Such a network will inevitably build up any recurring time pattern in its inputs, even in the presence of noise or other patterns. The network builds up all periodic patterns at all time scales simultaneously, and each periodic pattern builds up in the delay loop with the corresponding recurrence time.

Many random, periodically repeating binary sequences were presented to the network. The network behavior for an input pulse train that repeats the same 11-step sequence (. . . 10101100101 . . .) is shown in Figure 10a. Here coincidence windows are one timestep (sampling period). The plot on the right shows the value of the circulating pattern for each delay loop as a function of time. The evolution of the circulating pattern can be interpreted as the build up of a perceptual image. The network builds up the 11-step sequence first in the loop whose recurrence time is 11 steps, and later in the loop whose recurrence time is 22 time steps. This behavior is entirely understandable, since it is in the delay loop whose recurrence time equals the period of the repeating pattern that the previous pattern maximally coincides with its repetition. If pulses are randomly added to the pattern (noise), then their pulse patterns do not reliably recur at periodic intervals, and consequently do not build up in any one delay loop. The network detects periodic patterns and enhances them relative to aperiodic ones.

Such a network can separate out different, metrically-unrelated temporal patterns. One can combine multiple complex beat patterns with different periodicities and present them to the network. If there are multiple recurring periodic pattern, then each pattern builds up in the loop with the recurrence time matching its own period. Perhaps the most powerful feature of this behavior is that multiple patterns can be

separated and identified by examining the waveforms that circulate in these loops. In the loops where the patterns have built themselves up, the pulse pattern in each of the loops resembles one of the constituent patterns. There is thus a means of building up auditory objects, both rhythmic and tonal, out of recurrent temporal pattern invariances. The invariant nature of relations within objects and the constantly changing relations between objects permits these different objects to be separated in a relatively simple and elegant way.

Two kinds of stimuli were originally explored: the periodic binary sequences discussed above, and concurrent vowels with different fundamentals. The former were used to explore rhythm- and periodicity-detection, while the latter were used to study auditory object formation and stream segregation.

Several extensions of these recurrent nets have been implemented. First, the buildup rule has been modified. A straight multiplicative rule amplifies patterns geometrically, such that waveforms rapidly become distorted in their amplitudes (but not in their zero-crossing patterns). Secondly, while the networks handle isochronous rhythms well, they are less effective at using rhythmic expectations that have been built up when there are random delays inserted into sequences (pattern jitter). The second set of stimuli include concurrent, double vowels. Here the recurrent nets have been extended to receive input from a simulated auditory nerve front end. Each characteristic frequency channel has a full set of delay loops within which the patterns from that channel build up. Preliminary results suggest that recurrent timing networks can operate effectively on a frequency-by-frequency basis to segregate double vowels.

Recurrent timing nets for computing rhythmic expectation

Two rhythmic patterns have been analyzed using running autocorrelation and recurrent timing nets. The first is the beat pattern of La Marseillaise, encoded as a 64-step binary sequence, kindly provided by Bill Sethares. The second is a fragment of *Presto energetico from the Musica Recercata per pianoforte* (1951–53) by Gyorgy Ligeti, which was kindly provided by Marc Leman and Dirk Moelants. This is the same Ligeti fragment that was analyzed by a variety of methods in (Leman & Verbeke, 2000).

One of the shortcomings of the simple 5% buildup rule discussed above is that given a periodic signal, the response pattern builds up geometrically, and this dramatically favors shorter periodicities over longer ones. In order to rectify this imbalance, the buildup factor B that regulates the rate of increase of the loop signal was adjusted in proportion to the loop delay LD_k, i.e., $B_k = LD_k/100$, such that longer delay loops have proportionately higher facilitative factors. This equalizes in a crude way shorter and longer loops, which have differences in the number of times the signals are subjected to coincidence and facilitation per unit time. Subsequent applications of these networks to the problem of separating concurrent vowels have used buildup rules that saturate more gracefully, where the output of a given coincidence unit is the minimum of direct and circulating inputs plus some fraction of their difference. The rule that describes the coincidence operation was $S_k(t) = \min(S_{direct}(t), B * S_{direct}(t) * S_{loop}(t))$, where $S_k(t)$ is the ourput of coincidence element k associated with delay loop of recurrence time LD_k, $S_{direct}(t)$ is the incoming direct input signal, and $S_{loop}(t) = S_k(t - LD_k)$.

The response of the network to the La Marseillaise beat-pattern is shown in Figure 10b. The plot shows prominent sustained beat-periodicities at 16, 32, 64, 96, and 128 time steps, with the dominant periodicity being at the repetition period of the whole pattern (64) and its double (128). Recurrent nets simultaneously build up all of the meters and submeters that are consistently present in their inputs, and sometimes hidden, embedded sub-patterns were detected in the arbitrary repeating pulse sequences discussed above.

Recurrent timing networks can therefore be used to find embedded meters simply by examining the patterns in delay channels that grow past some detection threshold. Their pattern detection properties parallel those of the periodicity transform of Sethares (2001, this issue), which is similarly based on correlation. Both methods in their analysis of the input signal search the space of periodic patterns to find those periodicities present. The periodicity transform does this in a more directed and sequential way that eliminates redundant patterns by collapsing pattern multiples (e.g., the 64-step and 128-step periodicities in La Marseillaise are collapsed into the 64-step pattern). In contrast, the recurrent network performs its operations in parallel, and, like its autocorrelation cousin, retains all of the subpatterns along with multiples of repeating patterns. The periodicity transform is thus suited to finding a single dominant pattern, while the autocorrelation-like methods are more suited to presenting all of the possible (competing) patterns that might be heard out.

A more difficult test is the Ligeti fragment. The waveform, envelope, autocorrelogram, and recurrent network response for the Ligeti fragment are shown in Figure 11a–c. The running rms of the waveform of the audio recording of the piano peformance (a) was computed every 10 ms using a 50 ms moving window, and the whole rms waveform was rescaled to the range (0, 1). This low-frequency envelope of the waveform (b) was analyzed using running autocorrelation and a recurrent timing net.

The running autocorrelation (RAC) of the Ligeti fragment is shown in (c). The function is a signal expansion that uses no windowing (i.e., it has the temporal resolution of the signal itself): $RAC(tau, t) = X(t) * X(t - tau)$. The autocorrelogram plots the running autocorrelation, which depicts all periodicities (tau) as a function of time (t), making it is useful for analyzing time-varying signals. Autocorrelograms have been used to display running all-order population-interval distributions in response to time-varying complex stimuli (Cariani & Delgutte, 1996a) and to display the periodicity structure of music (Leman & Carreras, 1997). The running

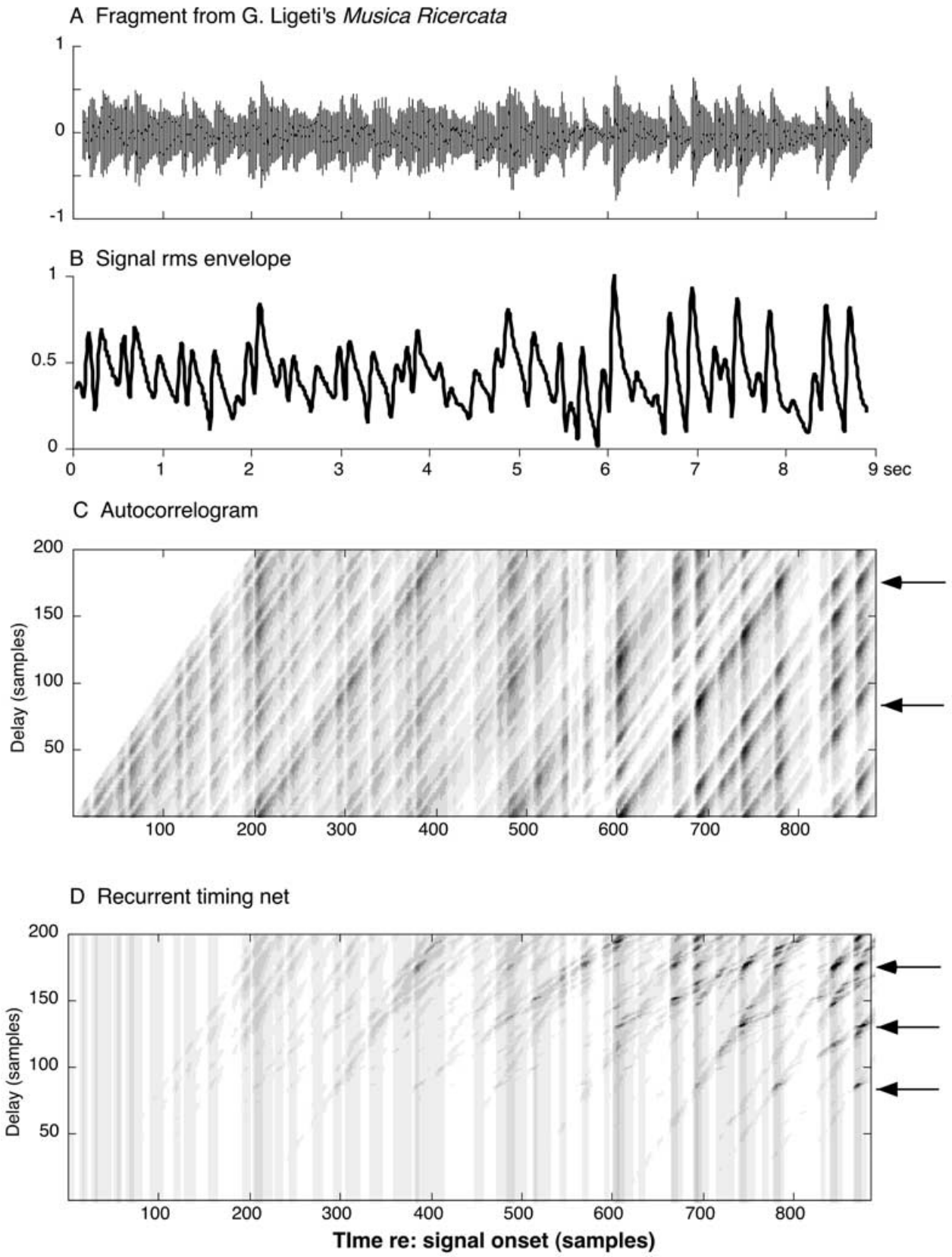

Fig. 11. Analysis of complex rhythmic pattern in music. (a) Waveform fragment from Ligeti, *Musica Ricercata.* (b) Rms envelope of the waveform. (c) Autocorrelogram (running autocorrelation) of the envelope. (d) Response of the recurrent timing net. Arrows indicate delay channels that built up prominent patterns.

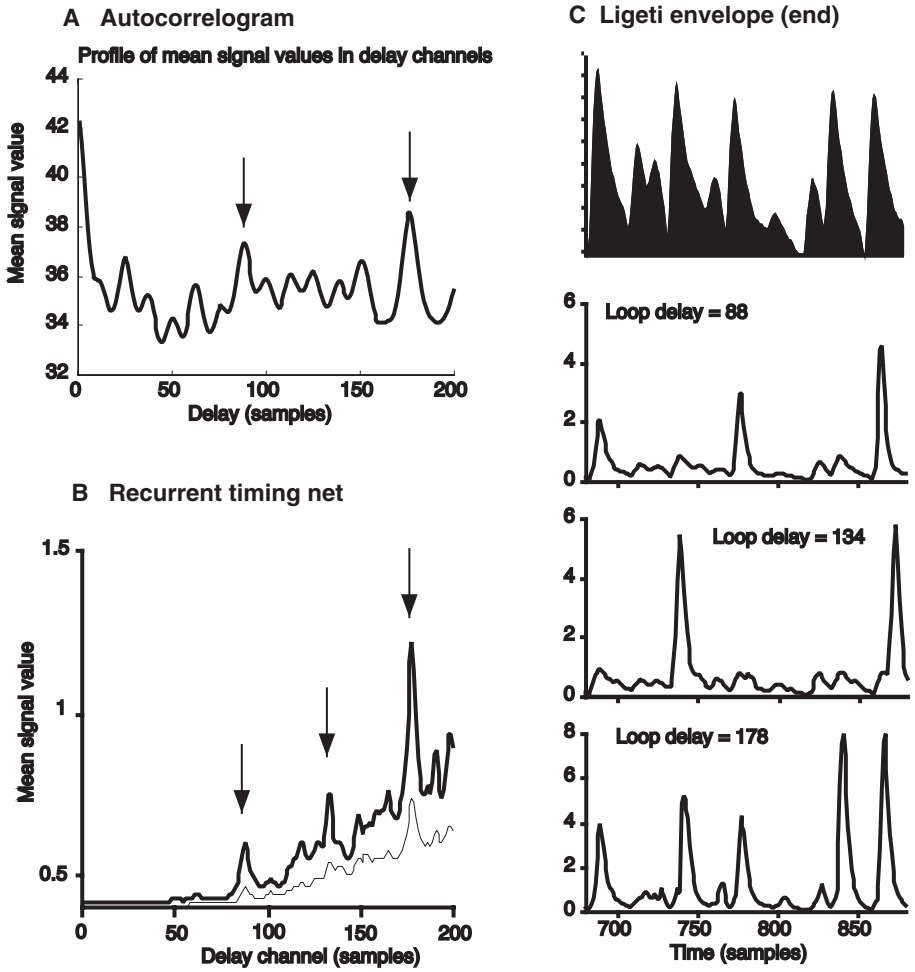

Fig. 12. Analysis of Ligeti fragment by autocorelograms and timing net. (a) Average autocorrelation value as a function of delay (mean value of autocorrelogram). (b) Average amplitude of signals in timing net loops as a function of loop delay. Thicker line shows average signal over the last 200 samples (2 s); thin line, over the whole fragment. (c) Top: End of Ligeti fragment. Below: Waveforms built up in the three most highly activated delay loops.

autocorrelation can be compared with other running displays of time struture: cochleograms (Lyon & Shamma, 1996; Slaney & Lyon, 1993), tempograms and strobed auditory images (Patterson et al., 1995). The running autocorrelogram also affords a method of periodicity detection. If temporal windowing is applied to the running autocorrelation at each delay lag, one can find dominant periodicities by comparing mean correlations as a function of lag (Figure 12a). Both mean and standard deviation profiles of the signals circulating through the delay channels indicate the presence of strong rhythmic patterns with durations at 88 and 178 samples (0.88 and 1.78 s). These are most apparent in the peaks in Figure 12a, and are indicated by arrows in the autocorrelogram of Figure 11c.

The response of the recurrent net to the Ligeti fragment is similarly shown in Figure 11d and in Figures 12b and c. The recurrent net builds up the same periodicities that are seen in the running autocorrelation: at 88 and 178 samples duration, plus an additional one at 134 samples (1.34 s). These can also be seen in the peaks in the mean signal values in each delay loop that are plotted in Figure 12b. The additional peak is due to the difference between the non-iterated autocorrelation of the autocorrelogram and the exponential nature of the iterated pattern-buildup process of the recurrent net. The upward trend in the mean signal strength profile is due to the loop-dependent scaling of the buildup factor that was discussed above.

Although one can read off the periods of dominant temporal patterns by examining mean signal amplitudes and variances, the main advantage of a timing net over an array of oscillators or periodicity detectors is that it builds up and retains the *form* of the periodic pattern with all of its metric microstructure (Moelants, 1997). The last two seconds of the Ligeti fragment are shown in Figure 12c (top plot). Below it are shown the waveforms for the signals that were circulating in the maximally activated delay loops. In the longest delay channel (178 samples, 1.78 s, bottom plot) is the whole repeating pattern, while rhythmic subpatterns at 88 samples

(1 : 2) and 134 samples (3 : 4) present themselves. The network thus represents the whole and its parts, such that either can be further analyzed.

The foregoing rhythmic examples are simple illustrations of the kinds of behaviors that recurrent timing nets exhibit, without any formal attempt to link these to patterns of rhythm perception or to empirically test these networks as psychoneural models. In order to do so, we would want to examine how many repetitions of a pattern are necessary to build up an expectation of a given strength, as well as how many beat omissions and how much pattern-jitter both human listeners and timing nets could tolerate. We would want to know whether human and neural network find the same patterns most salient.

At present timing nets function as broad heuristics for how the auditory system and the brain in general might process temporal patterns to form auditory objects and temporal expectations. We ponder whether mass temporal comparison operations could be realized interactions between ascending and descending fiber systems at the level of colliculus and thalamus. This could be accomplished using synaptic inputs or via axonal cross-talk. Certainly a great deal of further refinement will be necessary before these networks reach a stage where specific psychological and neurophysiological hypotheses can be empirically tested.

In many ways the goals and operation of recurrent timing networks are most similar to those of Patterson's strobed temporal integration architecture (Patterson et al., 1995). Both build up auditory images by comparing a signal with its immediate past. While Patterson's model uses an onset-triggered comparison process, these recurrent timing nets continuously compute with all loop delays, which yields a more systematic analysis of the signal.

In this paper, we have treated the processing of pitch and rhythm in a disjunctive way, using feedforward nets for pitch analysis and recurrent ones for rhythm. However, recurrent nets also form and separate objects by common pitch (Cariani, 2001d). Feedforward nets could potentially be used for extraction of rhythmic subpatterns and for tempo matching. The two qualities of pitch and rhythm, despite their very different tempi, have a surprising number of common perceptual properties (ratio relations, pattern fusions and separations) that lend themselves to similar computational strategies (autocorrelation, comb filters, oscillator nets, timing nets, modulation spectra). This suggests that pitch and rhythm (and still longer structures) might well be handled by similar temporal processing mechanisms, albeit operating on different time scales (Scheirer, 1998; Scheirer, 2000). Ideally, such a general mechanism should incorporate a means of forming objects and of analyzing their properties. Recurrent nets would first build up and stabilize auditory objects on the basis of temporal pattern coherence. This is a mechanism that is sensitive to onset asynchronies and abrupt changes in phase. Periodic waveforms would then be built up in delay loops whose pitch and timbral properties could be subsequently compared with other temporal patterns using feedforward networks that are largely indifferent to phase.

More broadly, these networks can also be seen as temporal versions of adaptive resonance networks (Grossberg, 1988), in which patterns are temporally rather than spatially coded. Adaptive resonance networks, of course, have a huge edifice of theory and practice developed over decades of experience that can potentially be transferred into time domain networks. In both adaptive resonance and recurrent timing networks, there is an interplay of incoming sensory data and central circulating patterns that makes for bottom-up/top-down codeterminations. We concentrate here on the dynamic formation of patterns rather than recognition of incoming patterns vis-à-vis stored pattern archetypes, but one can conceive of central neural assemblies that emit temporal patterns that then facilitate the buildup of like patterns if they are present in incoming sensory data. Periodicity-assimilating units (John, 1967; Morrell, 1967) as well as those that encode the expected timings of events have been observed in neurophysiological studies of conditioning. Thus far, the simple recurrent timing nets presented here do not exploit mismatches between incoming patterns and network expectations as they do in adaptive resonance circuits. One can foresee incorporation of temporally-precise inhibitory interactions that implement anti-coincidence operations that make detections of such mismatches possible in timing nets as well. Finally, adaptive resonance networks are adaptive – they alter their internal structure contingent on experience in order to improve performance – while the timing nets thus far developed are not. Here, too, the improvements that must be made are fairly straightforward, involving the incorporation of Hebbian rules that operate on temporal correlations and anticorrelations, e.g., the kinds of short-term synaptic facilitations that are now under active investigation. Perhaps the most exciting prospect is that delay loops could be formed on the fly even in randomly-connected nets by such short-term facilitations borne by temporal correlations. This would then mean that, once again, it is the stimulus that organizes the tissue, not only on the longer timescales of recovery to injury, but also, potentially, on the shorter timescales in which music impresses its temporal form on the brain.

Conclusions

We have explored some the implications that auditory temporal codes and neural timing nets might hold for music perception. In the realm of musical tonality, temporal microstructure in the form of autocorrelation-like population-interval distributions may be responsible for basic similarities between notes separated by simple frequency ratios. Pattern similarities between population-interval distributions produced by individual notes and chords parallel human judgments of how well particular notes fit in with particular chords. We have outlined how feedforward timing networks operating on such neural temporal representations might

compute such similarities, which result from the mass statistical behavior of intervals interacting in coincidence arrays. In the realm of rhythm perception, we have shown how recurrent timing nets can build up temporal expectations from recurring complex rhythmic patterns. Such networks provide an alternative to temporal processing based on clocks, oscillators, periodicity and duration detectors, and time hierarchies. Although, neural timing networks are presently at a rudimentary state of development and testing, they nevertheless bear promise as neurally-realizable models of musical perception. Temporal coding and neural timing nets potentially provide a unifying neurocomputational framework for music perception that encompasses pitch, timbre, rhythm, and still longer temporal patterns.

References

Bharucha, J.J. (1991). Pitch, harmony and neural nets: a psychological perspective. In: P. Todd & G. Loy (Eds.), *Connectionism and Music* (pp. 84–99). Cambridge: MIT Press.

Bharucha, J.J. (1999). Neural nets, temporal composites, and tonality. In: D. Deutsch (Ed.) (pp. 413–440). San Diego: Academic Press.

Boomsliter, P. & Creel, W. (1962). The long pattern hypothesis in harmony and hearing. *Journal of Music Theory*, *5*, 2–31.

Boring, E.G. (1942). *Sensation and Perception in the History of Experimental Psychology*. New York: Appleton-Century-Crofts.

Braitenberg, V. (1961). Functional interpretation of cerebellar histology. *Nature*, *190*, 539–540.

Brugge, J.F., Anderson, D.J., Hind, J.E., & Rose, J.E. (1969). Time structure of discharges in single auditory nerve fibers of the squirrel monkey in response to complex periodic sounds. *Journal of Neurophysiology*, *32*, 386–401.

Cariani, P. (1995). As if time really mattered: temporal strategies for neural coding of sensory information. *Communication and Cognition – Artificial Intelligence (CC-AI), 12*(1–2), 161–229. Reprinted in: K Pribram (Ed.), *Origins: Brain and Self-Organization*, Hillsdale, NJ: Lawrence Erlbaum, 1994; 1208–1252.

Cariani, P. (1999). Temporal coding of periodicity pitch in the auditory system: an overview. *Neural Plasticity*, *6*, 147–172.

Cariani, P. (2001a). Neural timing nets for auditory computation. In: S. Greenberg & M. Slaney (Eds.), *Computational Models of Auditory Function* (pp. 235–249). Amsterdam: IOS Press.

Cariani, P. (2001b). Symbols and dynamics in the brain. *Biosystems*, *60*, 59–83.

Cariani, P. (2001c). Temporal coding of sensory information in the brain. *Acoust. Sci. & Tech*., *22*, 77–84.

Cariani, P. (2001d). Neural timing nets. *Neural Networks*, *14*, 737–753.

Cariani, P., Delgutte, B., & Tramo, M. (1997). Neural representation of pitch through autocorrelation. *Proceedings, Audio Engineering Society Meeting (AES), New York, September 1997, Preprint #4583 (L-3).*

Cariani, P.A. & Delgutte, B. (1996a). Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *Journal of Neurophysiology*, *76*, 1698–1716.

Cariani, P.A. & Delgutte, B. (1996b). Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase-invariance, pitch circularity, and the dominance region for pitch. *Journal of Neurophysiology*, *76*, 1717–1734.

Cherry, C. (1961). Two ears – but one world. In: W.A. Rosenblith (Ed.), *Sensory Communication* (pp. 99–117). New York: MIT Press/John Wiley.

Chistovich, L.A. (1985). Central auditory processing of peripheral vowel spectra. *Journal of the Acoustical Society of America*, *77*, 789–805.

Chistovich, L.A. & Malinnikova, T.G. (1984). Processing and accumulation of spectrum shape information over the vowel duration. *Speech Communication*, *3*, 361–370.

Clarke, E.F. (1999). Rhythm and timing in music. In: D. Deutsch (Ed.) (pp. 473–500). San Diego: Academic Press.

Clynes, M. & Walker, J. (1982). Neurobiologic functions, rhythm, time, and pulse in music. In: M. Clynes (Ed.), *Music, Mind, and Brain: the Neuropsychology of Music* (pp. 171–216). New York: Plenum.

Cohen, M.A., Grossberg, S., & Wyse, L.L. (1994). A spectral network model of pitch perception. *J. Acoust. Soc. Am.*, *98*, 862–879.

de Boer, E. (1956). *On the "residue" in hearing.* Unpublished Ph.D., University of Amsterdam.

de Boer, E. (1976). On the "residue" and auditory pitch perception. In: W.D. Keidel & W.D. Neff (Eds.), *Handbook of Sensory Physiology* (Vol. 3, pp. 479–583). Berlin: Springer Verlag.

de Cheveigné, A. (1998). Cancellation model of pitch perception. *Journal of the Acoustical Society of America*, *103*, 1261–1271.

Delgutte, B. (1996). Physiological models for basic auditory percepts. In: H. Hawkins, T. McMullin, A.N. Popper, & R.R. Fay (Eds.), *Auditory Computation* (pp. 157–220). New York: Springer Verlag.

Demany, L. & Semal, C. (1988). Dichotic fusion of two tones one octave apart: evidence for internal octave templates. *Journal of the Acoustical Society of America*, *83*, 687–695.

Demany, L. & Semal, C. (1990). Harmonic and melodic octave templates. *Journal of the Acoustical Society of America*, *88*, 2126–2135.

DeWitt, L.A. & Crowder, R.G. (1987). Tonal fusion of consonant musical intervals: The oomph in Stumpf. *Perception and Psychophysics*, *41*, 73–84.

Epstein, D. (1995). *Shaping Time: Music, Brain, and Performance.* New York: Simon & Schuster Macmillan.

Essens, P.J. & Povel, D.-J. (1985). Metrical and nonmetrical representations of temporal patterns. *Perception and Psychophysics*, *37*, 1–7.

Evans, E.F. (1978). Place and time coding of frequency in the peripheral auditory system: some physiological pros and cons. *Audiology*, *17*, 369–420.

Fay, R.R. (1988). *Hearing in vertebrates: a psychophysics databook*. Winnetka, Ill.: Hill-Fay Associates.

Fraisse, P. (1978). Time and rhythm perception. In: E.C. Carterette & M.P. Friedman (Eds.), *Handbook of Perception. Volume VIII. Perceptual Coding* (pp. 203–254). New York: Academic Press.

Goldstein, J.L. (1973). An optimum processor theory for the central formation of the pitch of complex tones. *Journal of the Acoustical Society of America*, *54*, 1496–1516.

Goldstein, J.L. & Srulovicz, P. (1977). Auditory-nerve spike intervals as an adequate basis for aural frequency measurement. In: E.F. Evans & J.P. Wilson (Eds.), *Psychophysics and Physiology of Hearing* (pp. 337–346). London: Academic Press.

Gray, P.M., Krause, B., Atema, J., Payne, R., Krumhansl, C., & Baptista, L. (2001). The music of nature and the nature of music. *Science*, *291*, 52–54.

Greenberg, S. (1980). *Neural Temporal Coding of Pitch and Vowel Quality: Human Frequency-Following Response Studies of Complex Signals*. Los Angeles: UCLA Working Papers in Phonetics #52.

Grossberg, S. (1988). *The Adaptive Brain, Vols I. and II*. New York: Elsevier.

Hall III, J.W. & Peters, R.W. (1981). Pitch for nonsimultaneous successive harmonics in quiet and noise. *Journal of the Acoustical Society of America*, *69*, 509–513.

Handel, S. (1989). *Listening*. Cambridge: MIT Press.

Hebb, D.O. (1949). *The Organization of Behavior*. New York: Simon & Schuster.

Hindemith, P. (1945). *The Craft of Musical Composition. I. Theoretical Part* (Arthur Mendel, Trans.). New York: Associated Music Publishers.

Hunt, F.V. (1978). *Origins in Acoustics*. Woodbury, NY: Acoustical Society of America.

Jeffress, L.A. (1948). A place theory of sound localization. *J. Comp. Physiol. Psychol.*, *41*, 35–39.

John, E.R. (1967). Electrophysiological studies of conditioning. In: G.C. Quarton, T. Melnechuk, & F.O. Schmitt (Eds.), *The Neurosciences: A Study Program* (pp. 690–704). New York: Rockefeller University Press.

John, E.R. (1972). Switchboard vs. statistical theories of learning and memory. *Science*, *177*, 850–864.

Jones, M.R. (1976). Time, our lost dimension: toward a new theory of perception, attention, and memory. *Psychological Review*, *83*, 323–255.

Jones, M.R. (1978). Auditory patterns: studies in the perception of structure. In: E.C. Carterette & M.P. Friedman (Eds.), *Handbook of Perception. Volume VIII. Perceptual Coding* (pp. 255–288). New York: Academic Press.

Jones, M.R. (1987). Perspectives on musical time. In: A. Gabrielsson (Ed.), *Action and Perception in Rhythm and Music* (pp. 153–175). Stockholm: Royal Swedish Academy of Music.

Jones, M.R. & Hahn, J. (1986). Invariants in sound. In: V. McCabe & G.J. Balzano (Eds.), *Event Cognition: An Ecological Perspective* (pp. 197–215). Hillsdale, New Jersey: Lawrence Erlbaum.

Kaernbach, C. & Demany, L. (1998). Psychophysical evidence against the autocorrelation theory of auditory temporal processing. *Journal of the Acoustical Society of America*, *104*, 2298–2306.

Keidel, W. (1984). The sensory detection of vibrations. In: W.W. Dawson & J.M. Enoch (Eds.), *Foundations of Sensory Science* (pp. 465–512). Berlin: Springer-Verlag.

Keidel, W.D. (1992). The phenomenon of hearing: an interdiscplinary discussion. II. *Naturwissenschaften*, *79*, 347–357.

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

Krumhansl, C.L. (1990). *Cognitive Foundations of Musical Pitch*. New York: Oxford University Press.

Langner, G. (1983). Evidence for neuronal periodicity detection in the auditory system of the Guinea Fowl: implications for pitch analysis in the time domain. Experimental Brain Research, *52*, 33–355.

Langner, G. (1992). Periodicity coding in the auditory system. *Hearing Research*, *60*, 115–142.

Large, E.W. (1994). *Dynamic representation of musical structure.* Unpublished Ph.D., The Ohio State University.

Leman, M. (1999). Time domain filter model of tonal induction. *Tonality Induction, Proceedings of the Ninth FWO Research Society on Foundations of Music Research, University of Ghent, April 6–9*, *1999*, 53–87.

Leman, M. (2000). An auditory model of the role of short-term memory in probe-tone ratings. *Music Perception*, *17*, 481–510.

Leman, M. & Carreras, F. (1997). Schema and Gestalt: Testing the hypothesis if psychoneural isomorphism by computer simulation. In: M. Leman (Ed.), *Music, Gestalt, and Computing* (pp. 144–165). Berlin: Springer.

Leman, M. & Schneider, A. (1997). Origin and nature of cognitive and systematic musicology: an introduction. In: M. Leman (Ed.), *Music, Gestalt, and Computing* (pp. 11–29). Berlin: Springer.

Leman, M. & Verbeke, B. (2000). The concept of minimal 'energy' change (MEC) in relation to Fourier transform, auto-correlation, wavelets, AMDF, and brain-like timing networks – application to the recognition of repetitive rhythmical patterns in acoustic musical signals. In: K. Jokien, D. Helylen, & A. Nijholt (Eds.), *Proceedings, Workshop on Internalizing Knowledge (Nov. 22–24, 2000, Ieper, Belgium)* (pp. 191–200). Ieper, Belgium: Cele-Twente.

Licklider, J.C.R. (1951). A duplex theory of pitch perception. *Experientia*, *VII*(4), 128–134.

Licklider, J.C.R. (1954). "Periodicity" pitch and "place" pitch. *Journal of the Acoustical Society of America*, *26*, 945.

Licklider, J.C.R. (1956). Auditory frequency analysis. In: C. Cherry (Ed.), *Information Theory* (pp. 253–268). London: Butterworth.

Licklider, J.C.R. (1959). Three auditory theories. In: S. Koch (Ed.), *Psychology: A Study of a Science. Study I. Conceptual and Systematic* (Vol. Volume I. Sensory, Perceptual, and Physiological Formulations, pp. 41–144). New York: McGraw-Hill.

Longuet-Higgins, H.C. (1987). *Mental processes: Studies in cognitive science*. Cambridge, Mass.: The MIT Press.

Longuet-Higgins, H.C. (1989). A mechanism for the storage of temporal correlations. In: R. Durbin, C. Miall, & G. Mitchison (Eds.), *The computing neuron* (pp. 99–104). Wokingham, England: Addison-Wesley.

Lyon, R. & Shamma, S. (1996). Auditory representations of timbre and pitch. In: H. Hawkins, T. McMullin, A.N. Popper, & R.R. Fay (Eds.), *Auditory computation* (pp. 221–270). New York: Springer Verlag.

Mach, E. (1898). *Popular scientific lectures, Third Edition*. La Salle, Illinois: Open Court.

MacKay, D.M. (1962). Self-organization in the time domain. In: M.C. Yovitts, G.T. Jacobi, & G.D. Goldstein (Eds.), *Self-Organizing Systems 1962* (pp. 37–48). Washington, D.C.: Spartan Books.

Makeig, S. (1982). Affective versus analytic perception of musical intervals. In: M. Clynes (Ed.), *Music, mind, and brain: The neuropsychology of music* (pp. 227–250). New York: Plenum.

McCulloch, W.S. (1947). Modes of functional organization of the cerebral cortex. *Federation Proceedings*, *6*, 448–452.

McCulloch, W.S. (1969). Of digital oscillators. In: K.N. Leibovic (Ed.), *Information processing in the nervous system* (pp. 293–296). New York: Springer Verlag.

McKinney, M. (1999). A possible neurophysiological basis of the octave enlargement effect. *Journal of the Acoustical Society of America*, *106*, 2679–2692.

Meddis, R. & Hewitt, M.J. (1991a). Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I. Pitch identification. *Journal of the Acoustical Society of America*, *89*, 2866–2882.

Meddis, R. & Hewitt, M.J. (1991b). Virtual pitch and phase sensitivity of a computer model of the auditory periphery. II. Phase sensitivity. *Journal of the Acoustical Society of America*, *89*, 2883–2894.

Meddis, R. & O'Mard, L. (1997). A unitary model of pitch perception. *Journal of the Acoustical Society of America*, *102*, 1811–1820.

Meyer, L.B. (1956). *Emotion and meaning in music*. Chicago: University of Chicago.

Moelants, D. (1997). A framework for the subsymbolic description of meter. In: M. Leman (Ed.), *Music, Gestalt, and Computing* (pp. 263–276). Berlin: Springer.

Moore, B.C.J. (1997a). *Introduction to the psychology of hearing* (Fourth ed.). London: Academic Press.

Moore, B.C.J. (1997b). *Introduction to the psychology of hearing* (Fourth ed.). London: Academic Press.

Morrell, F. (1967). Electrical signs of sensory coding. In: G.C. Quarton, T. Melnechuck, & F.O. Schmitt (Eds.), *The neurosciences: A study program* (pp. 452–469). New York: Rockefeller University Press.

Ohgushi, K. (1983). The origin of tonality and a possible explanation for the octave enlargement phenomenon. *Journal of the Acoustical Society of America*, *73*, 1694–1700.

Orbach, J. (1998). *The neuropsychological theories of Lashley and Hebb*. Lanham: University Press of America.

Palmer, A.R. (1992). Segregation of the responses to paired vowels in the auditory nerve of the guinea pig using autocorrelation. In: M.E.H. Schouten (Ed.), *The auditory processing of speech* (pp. 115–124). Berlin: Mouton de Gruyter.

Parncutt, R. (1989). *Harmony: A psychoacoustical approach*. Berlin: Springer-Verlag.

Patterson, R.D. (1986). Spiral detection of periodicity and the spiral form of musical scales. *Psychology of Music*, *14*, 44–61.

Patterson, R.D. (1994). The sound of a sinusoid: time-interval models. *Journal of the Acoustical Society of America*, *96*, 1560–1586.

Patterson, R.D., Allerhand, M.H., & Giguere, C. (1995). Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform. *Journal of the Acoustical Society of America*, *98*, 1890–1894.

Perkell, D.H. & Bullock, T.H. (1968). Neural coding. *Neurosciences Research Program Bulletin*, *6*, 221–348.

Pressnitzer, D., Patterson, R.D., & Krumboltz, K. (2001). The lower limit of melodic pitch. *Journal of the Acoustical Society of America*, *109*, 2074–2084.

Rieke, F., Warland, D., de Ruyter van Steveninck, R., & Bialek, W. (1997). *Spikes: Exploring the neural code*. Cambridge: MIT Press.

Rose, J.E. (1980). Neural correlates of some psychoacoustical experiences. In: D. McFadden (Ed.), *Neural mechanisms of behavior* (pp. 1–33). New York: Springer Verlag.

Scheirer, E.D. (1998). Tempo and beat analysis of acoustic musical signals. *Journal of the Acoustical Society of America*, *103*, 588–601.

Scheirer, E.D. (2000). *Music listening systems.* Cambridge, Mass.: Ph.D. Dissertation, M.I.T.

Schneider, A. (1997). "Verschmelzung", tonal fusion, and consoance: Carl Stumpf revisited. In: M. Leman (Ed.), *Music, Gestalt, and Computing*. Berlin: Springer.

Schneider, A. (2001, in press). Inharmonic sounds: implications as to "pitch", "timbre" and "consonance". *Journal of New Music Research*.

Schouten, J.F. (1940). The residue, a new concept in subjective sound. Proceedings Koninklijke Nederlandse van Wete-Schappen, *43*, 356–365.

Schouten, J.F., Ritsma, R.J., & Cardozo, B.L. (1962). Pitch of the residue. *Journal of the Acoustical Society of America*, *34*, 1418–1424.

Schreiner, C.E. & Langner, G. (1988). Coding of temporal patterns in the central auditory system. In: G.M. Edelman, W.E. Gall, & W.M. Cowan (Eds.), *Auditory function: Neurobiological bases of hearing* (pp. 337–362). Wiley: New York.

Schwarz, D.W.F. & Tomlinson, R.W.W. (1990). Spectral response patterns of auditory cortical neurons to harmonic complex tones in alert monkey (Macaca mulatta). *Journal of Neurophysiology*, *64*, 282–298.

Seneff, S. (1985). *Pitch and spectral analysis of speech based on an auditory synchrony model.* Unpublished Ph.D., M.I.T.

Seneff, S. (1988). A joint synchrony/mean-rate model of auditory speech processing. *Journal of Phonetics*, *16*, 55–76.

Sethares, W.A. (1999). *Tuning, timbre, spectrum, scale*. London: Springer.

Sethares, W.A. & Staley, T. (2001). Meter and periodicity in musical performance. *Journal of New Music Research*, *30*, 149–158.

Shamma, S. & Sutton, D. (2000). The case of the missing pitch templates: How harmonic templates emerge in the early auditory system. *Journal of the Acoustical Society of America*, *107*, 2631–2644.

Shepard, R.N. (1964). Circularity in judgments of relative pitch. *Journal of the Acoustical Society of America*, *36*, 2346–2353.

Siebert, W.M. (1968). Stimulus transformations in the peripheral auditory system. In: P.A. Kollers & M. Eden (Eds.), *Recognizing patterns* (pp. 104–133). Cambridge: MIT Press.

Simmons, A.M. & Ferragamo, M. (1993). Periodicity extraction in the anuran auditory nerve. *J. Comp. Physiol. A*, *172*, 57–69.

Slaney, M. (1998). Connecting Correlograms to neurophysiology and psychoacoustics. In: A.R. Palmer, A. Rees, A.Q. Summerfield, & R. Meddis (Eds.), *Psychophysical and physiological advances in hearing* (pp. 563–569). London: Whurr.

Slaney, M. & Lyon, R.F. (1993). On the importance of time – a temporal representation of sound. In: M. Cooke, S. Beet, & M. Crawford (Eds.), *Visual representations of speech signals* (pp. 95–118). New York: John Wiley.

Terhardt, E. (1973). Pitch, consonance, and harmony. *Journal of the Acoustical Society of America*, *55*, 1061–1069.

Terhardt, E. (1979). Calculating virtual pitch. *Hearing Research*, *1*, 155–182.

Thatcher, R.W. & John, E.R. (1977). *Functional neuroscience, Vol. I. Foundations of cognitive processes*. Hillsdale, NJ: Lawrence Erlbaum.

Tramo, M.J. (2001). Music of the hemispheres. *Science*, *291*, 54–56.

Tramo, M.J., Cariani, P.A. Delgutte, B., & Braida, L.D. (2001). Neurobiological foundations for the theory of harmony in Western tonal music. *Annals of the New York Academy of Sciences*, *930*, 92–116.

Troland, L.T. (1929a). *The principles of psychophysiology, Vols I–III*. New York: D. Van Nostrand.

Troland, L.T. (1929b). The psychophysiology of auditory qualities and attributes. *Journal of Geu. Psychol.*, *2*, 28–58.

Turgeon, M. (1999). *Cross-spectral auditory grouping using the paradigm of rhythmic masking release (Ph.D. Thesis)*. Montreal: McGill University.

Turgeon, M., Bregman, A.S., & Ahad, P.A. (in press). Rhythmic masking release: Contribution of cues for perceptual organization to the cross-spectral fusion of concurrent narrow-band noise. *Journal of the Acoustical Society of America*.

van Noorden, L. (1982). Two channel pitch perception. In: M. Clynes (Ed.), *Music, mind and brain* (pp. 251–269). New York: Plenum.

Weinberg, N.M. (1999). Music and the auditory system. In: D. Deutsch (Ed.) (pp. 47–112). San Diego: Academic Press.

Wever, E.G. (1949). *Theory of hearing*. New York: Wiley.

White, L.J. & Plack, C.J. (1998). Temporal processing of the pitch of complex tones. *Journal of the Acoustical Society of America*, *103*(4), 2051–2059.

Wiegrebe, L. (2001). Searching for the time constant of neural pitch extraction. *Journal of the Acoustical Society of America*, *109*(3), 1082–1091.

A TEMPORAL MODEL FOR PITCH MULTIPLICITY AND TONAL CONSONANCE

Peter Cariani

Eaton Peabody Laboratory of Auditory Physiology, Mass. Eye & Ear Infirmary,
243 Charles St., Boston MA 02114 USA. Email: cariani@epl.meei.harvard.edu; www.cariani.com

ABSTRACT

One aspect of tonal consonance relevant to the pitch stability of chords and harmonic tension concerns the degree to which a collection of notes produces a unified, strong low pitch. We used a neurophysiologically-grounded temporal model for pitch to test Stumpf's general hypothesis that consonance is based on degree of tonal fusion. Patterns of consonance estimates generated from the model were compared with corresponding human judgments of consonance ("clearness") obtained by Kameoka & Kurogawa (1969).

The model used an auditory nerve simulation (72 fibers, CFs, 80-16000 Hz) to generate neural responses to pairs of tones separated by different musical intervals. The simulation used stimuli identical to those used in psychophysical experiments (pairs of pure tones and harmonic complexes, $F0_{root}$ = 440 Hz, n=1-6). Population-wide interspike interval distributions were then compiled from the responses of individual simulated auditory nerve fibers. The relative strengths of pitch-related interval patterns in these distributions were then measured using subharmonic sieves to yield predictions of which pitche(s) should be heard as well as their relative strengths (saliences). The estimated salience of the strongest pitch (maximum salience) is used as an estimate of pitch fusion and perceived consonance.

Consonance values estimated from the model were then compared with those obtained by K&K. The high degree of correspondences observed (r = 0.88 for "consonances" of pure tones and r = -0.97 for "absolute dissonances" of complex tones) support the viability of pitch-based models for consonance that depend on the strength of the fundamental bass.

1. BACKGROUND

The perception of concurrent and successive combinations of tones is central to psychological theories of musical harmony and melody. Consonance encompasses qualities that are created when tones are sounded together: from sensations of roughness to perceptions of unified, fused wholes (Tramo et al., 2001). Far from simple dependence on acoustics, perceived consonance is influenced by cochlear and neuronally-mediated processes that unfold on at least three different time scales. "Sensory" aspects of consonance are often used to describe perceived tone qualities of concurrent sounds presented in isolation. These aspects of consonance are likely to be mediated by cochlear filtering and the neuronal mechanisms responsible for perception of pitch and roughness. Contextual effects can arise when tones are preceded by other tones, and plausibly involve interactions between neuronal activity patterns associated with incoming sounds and activity patterns produced by sounds in the immediate past (~200 ms to seconds). These short-term contextual effects likely involve the same central neuronal mechanisms associated with echoic and working memory that mediate tonality induction (Leman, 2000) and melodic coherence (Deutsch, 1999; Snyder, 2000). Longer range contextual effects, such as particular patterns of tonal expectancies, almost inevitably involve experientially-acquired, (culturally) conditioned predispositions (hedonic and otherwise) that are mediated by long term memory (Bharucha, 1994).

Perception of consonance in music thus is thought involve "sensory" and "perceptual" processes that are relatively independent of context, as well as "cognitive" processes that (can) depend on musical context, both immediate and distant. We focus here on those "sensory" and "perceptual" aspects of consonance that apply to multiple concurrent tones presented in isolation. Models of consonance that are based on the perception of roughness examine the degree to which nearby harmonics beat (Helmholtz, 1885; Kameoka & Kuriyagawa, 1969a, 1969b; Plomp & Levelt, 1965; Sethares, 1999; Terhardt, 1974a), while pitch-based theories examine the strengths of virtual pitches created (Terhardt et al., 1982) and the degrees to which they fuse together (DeWitt & Crowder, 1987; Green & Butler, 2002; Lipps, 1905; Schneider, 1997).

2. AIMS

Our general aims here parallel those of others (Helmholtz, 1885; Parncutt, 1989; Terhardt, 1974b) who have sought to determine which aspects of musical perception are direct consequences of the physiological mechanisms that subserve auditory perception. Like others before us (Boomsliter & Creel, 1962; Licklider, 1951; Patterson, 1986; Rose, 1980), we also seek to ground pitch-based theories of tonal consonance in interspike interval representations. Recently global, population-wide interval statistics have been used to explain a wide range of musical pitch phenomena (Cariani, 2002; Cariani & Delgutte, 1996a, 1996b; Leman, 2000; Leman & Carreras, 1997; Meddis & Hewitt, 1991; Meddis & O'Mard, 1997; Slaney & Lyon, 1991). Here we present consonance predictions based on computer simulations of auditory nerve responses to dyads of pure and complex tones and compare them to consonance judgments obtained in psychophysical experiments (Kameoka & Kuriyagawa, 1969a, 1969b).

Our ultimate goal is to apply such models for pitch multiplicity to predict the pitch stability of chords so as to directly link low-level physiologically-grounded neuronal representations of pitch with

high-level music-theoretic notions of harmonic tension and relaxation. In order to achieve this ultimate goal, models need to incorporate effects of musical context, as embodied by neural processing operations that access neural representations held in both short and long-term memory stores (Bharucha, 1994; Leman, 2000). Such processes would lead to neuronal representations of music that dynamically unfold in time, combining successive presents with successive pasts.

3. METHOD

Auditory nerve model

Auditory nerve responses to dyads consisting of pure and complex tones were simulated in a computer model (Cariani, 2002). Ensembles of 72 auditory nerve fibers (ANFs) were simulated (3 fibers/CF), the 24 characteristic frequencies (CFs) being logarithmically-spaced between 80-16000 Hz. Single fiber responses were modeled using a series of relatively simple operations that parallel basic cochlear and neural processes: middle-ear filtering (attenuation: -20 dB < 100 Hz; 0 dB: 1-2 kHz; -20 dB < 10 kHz), cochlear band-pass frequency tuning (Bekesy-like gammatone filters; (Hartmann, 1998), pp. 254-256, filter order = 3, decay const. b = CF/5), hair cell transduction (half-wave rectification), synaptic filtering (low pass filter, rolloff > 2 kHz), neural rate-level functions, and uncorrelated, endogenous, "spontaneous" activity (Poisson process governed by driven - spontaneous rate (SR)). Three fibers representing three spontaneous rate/threshold classes (low, medium, and high SR) were simulated for each CF. An adaptive gain control similar to that of (Geisler & Greenberg, 1986) scales instantaneous firing rates according to the average input rms over the previous 4 ms so as to replicate rate compression and the quasi-linear character of fast amplitude modulations. Model parameters have been chosen that best replicate population-wide interspike interval distributions (PIDs) compiled from cat auditory nerve data (Cariani & Delgutte, 1996a). The model captures the response characteristics most relevant to temporal coding and auditory form perception: 1) the broadly-tuned nature of auditory nerve fiber responses at moderate-to-high levels, 2) faithful replication of the level-invariant forms of period histograms (Rose et al., 1971), impulse (click) responses, and threshold tuning curves (Kiang et al., 1965), and 3) rate saturation and the general spread of excitation at higher stimulus levels (Kim & Molnar, 1979). In addition, the model shows two-tone rate and synchrony suppression.

Stimuli

Waveforms were synthesized to match stimuli used in specific psychophysical experiments of Kameoka and Kuriyagawa, i.e. (Kameoka & Kuriyagawa, 1969a), fig. 5 and (Kameoka & Kuriyagawa, 1969b), fig. 7. Stimuli of 500 ms duration (44.1 kHz samples/s) were simulated at the specified levels: pure tone dyads at 57 dB SPL per component and complex tone dyads (harmonics 1-6) at 41-57 dB SPL per component.

Interspike interval distributions

The ANF model produces an array of instantaneous-firing-rate time series (post-stimulus time histograms, PSTHs). All-order interspike interval distributions from each fiber were estimated by computing the autocorrelation of the PSTHs (binwith = 0.1 ms). Interval distributions were weighted by SR fiber class and a human CF-distribution to form a population-interval distribution (PID) that serves as an estimate of the global all-order interval statistics of the human auditory nerve (Cariani & Delgutte, 1996a).

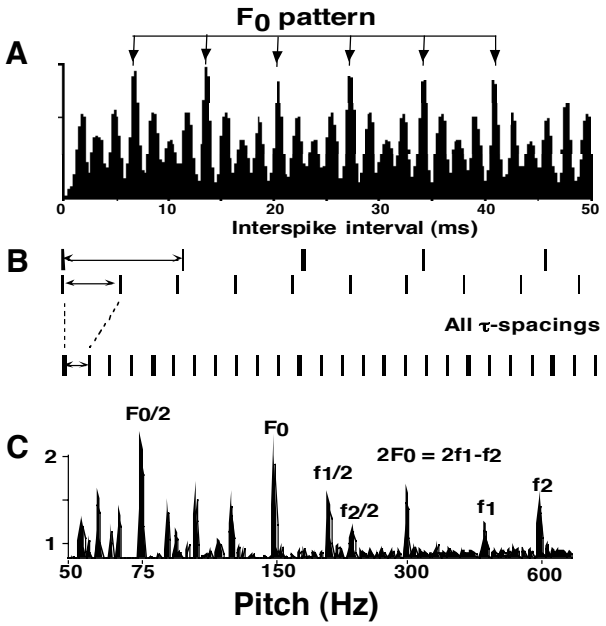

Figure 1. Pitch multiplicity and salience estimation. A. PID in response to two pure tones a fourth apart, 440 & 587 Hz. B. Representative sub-harmonic interval sieves for estimating pattern-saliences. C. Pitch map: observed distribution of pattern salience values. Exponential interval weighting window is not shown.

Pitch salience estimation

Our working hypothesis is that the neural correlate of a periodicity pitch percept involves a pattern of interspike intervals (e.g. intervals at the pitch period (1/F0) and its multiples (n/F0, the "undertone" series) rather than the relative number of intervals at only the pitch period itself. We want to estimate the relative pattern-strengths of different sets of interspike intervals that are associated with different pitches. The estimation procedure consists of two operations, the weighting of interspike intervals by their respective durations and application of periodic sieve that measures relative numbers of intervals associated with each pitch (Fig 1).

Interval weighting

Although we desire to include intervals at multiples of pitch periods in our analysis, a limit on the maximum interval duration that is processed in a central autocorrelation analyzer is needed to

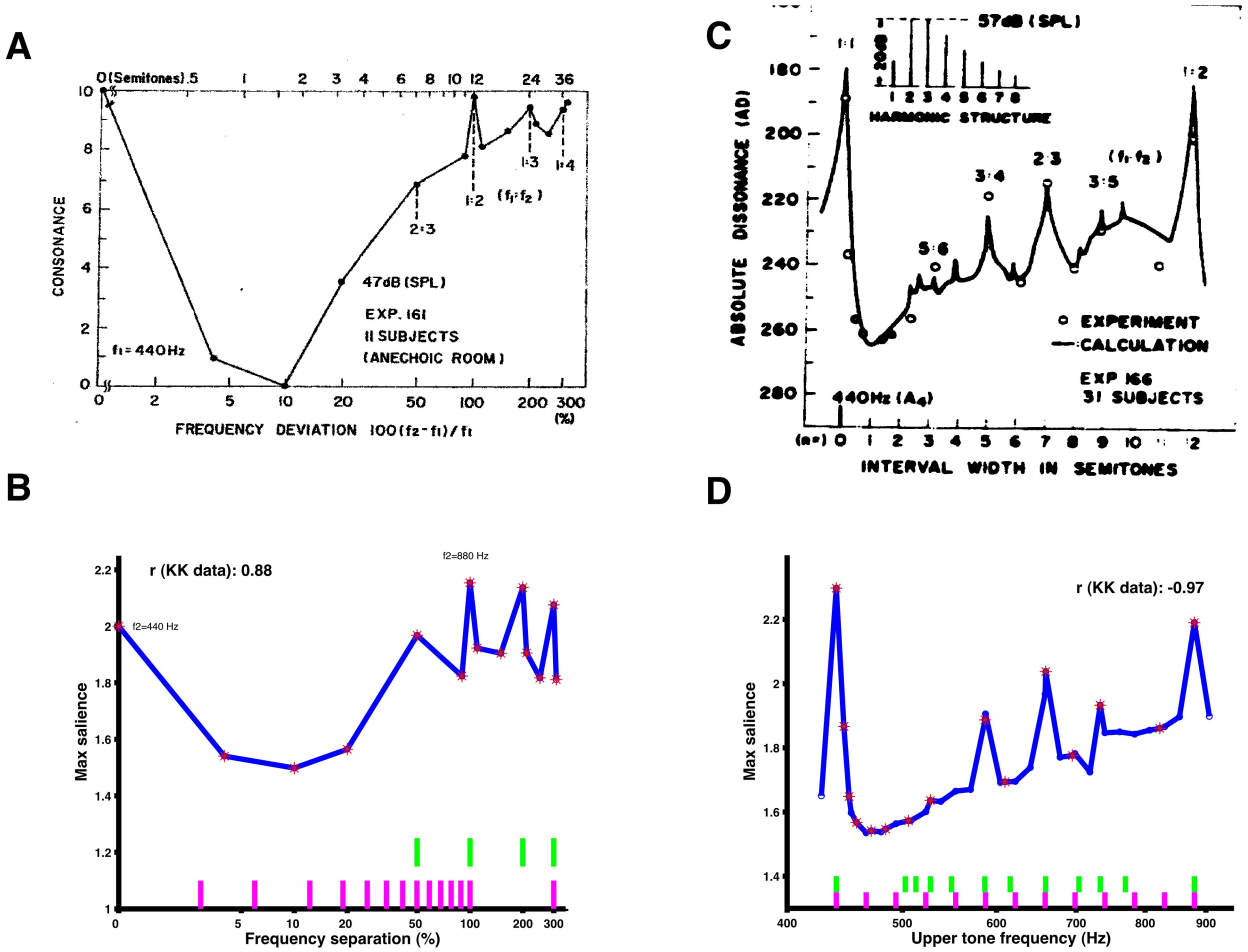

Figure 2. Comparisons of simulation-based consonance estimates with psychophysical observations. A. Subjective ratings of consonance ("clearness" vs. "tubidity") of pure tone dyads with different frequency separations in 11 human subjects, Fig. 5 in (Kameoka & Kuriyagawa, 1969a). B. Maximum pitch salience values for the same pure tone dyads computed from simulated auditory nerve population-interval distributions. Lower ticks: semitone intervals; upper ticks: just intervals (3:2, 2:1, 3:1, 4:1) C. Subjective ratings of absolute dissonance of complex tone dyads in 21 human subjects, Fig. 7 in (Kameoka & Kuriyagawa, 1969b). D. Maximum pitch salience values for complex tones (harmonics 1-6) computed from simulations as a function of the F0 of the upper tone (root F0 always 440 Hz).

account for the lower limit of musical pitch (~ 30 Hz, (Pressnitzer et al., 2001)), and the limited frequency resolution of the system. Discounting the influence of successively longer intervals achieves these effects. A neurocomputational interpretation would be that the interval analysis mechanism that subserves pitch has mostly only short time delays (< 33 ms) at its disposal. Simulations experiments with linear- and exponentially-tapering weights that decline with interval duration have suggested that an exponential decay with a time constant tau of ~9 ms best models the decline of pitch salience for periodicities below 100 Hz. The sharper the decay, the lower the saliences of periodicities < 100 Hz. The present model used CF-dependent taus: 100 Hz (.03s), 440 Hz (.016s), 880 Hz (.012s), 1320 Hz (.01s), and above (.009s).

Interval sieve

A dense set of periodic sieves (25-1000 Hz in 2 Hz steps) was applied to duration-weighted interval histograms in order to measure the relative pattern-strength of interval patterns associated with different perceived pitches. Sieves counted all bins that overlapped with the 100 us sieve holes. The salience of a particular pitch is estimated by dividing the mean density (intervals/bin) of pitch-related intervals in sieve-bins by the mean density of the whole distribution (background).

Maximum pitch salience

The pitch with the maximum salience value amongst all possible pitches is predicted to be the strongest pitch present and the one is most likely to be heard (cf. (Parncutt, 1989)).

The salience of this pitch produced by a given dyad or triad is one measure of the degree to which one pitch predominates over all other unrelated pitches. Harmonically-related pitches share intervals at their common subharmonics (nonexclusive allocation of intervals), so that their respective interval patterns interfere the least vis-à-vis the salience measure, while unrelated pitches reduce the salience of each other (because interval peaks in one raise the mean (background) density for the other). While the maximum pitch salience measure has its limitations, it captures many of the qualitative properties of "pitch fusion"(DeWitt & Crowder, 1987), "pitch unity", "tonal unity" (Lipps, 1905), fundamental bass, and virtual pitch strength (Parncutt, 1989).

4. RESULTS

Maximum saliences observed for pure and complex tone dyads are shown in Figure 2. For both pure and complex tones, maximal salience is highest for unison and the octave separations and lowest for separations near 1 semitone. In both cases, there was strong correspondence between psychophysical consonance ("clearness") and "absolute dissonance" ratings experimentally obtained by Kameoka and Kuriyagawa: $r = 0.88$ for consonances of pure tones and $r = -0.97$ for absolute dissonances of complex tones ((Kameoka & Kuriyagawa, 1969b), Fig. 7). In most cases, our population-interval model predictions followed those of K & K's roughness-based model for complex tones (C, solid lines, vs. D).

5. CONCLUSIONS

Representations based on all-order interspike intervals contain both overtone and undertone series, and in most respects their analysis parallels spectral operations based on coinciding harmonics and subharmonics (Terhardt et al., 1982). The present simulation demonstrates that interval-based models of low, "virtual" pitch can plausibly account for the consonance of pairs of pure and complex tones that are presented in isolation (without preceding tonal context). Since there exists a close relationship between frequency separations that give rise to maximal dissonance and critical band phenomena (Plomp & Levelt, 1965), these pitch fusion models may also predict critical bandwidths. It thus remains to be seen whether a neural basis for critical band phenomena might lie in the properties of population-wide interspike interval distributions of the auditory nerve (whose forms are partially determined by the properties of cochlear filters). Further studies will explore properties of population-interval distributions to model pitch multiplicities, stabilities, and similarities of triads along theoretical lines parallel to those of Parncutt (Parncutt, 1989) and Krumhansl (Krumhansl, 1990).

[This work was supported in part by NSF-EIA-BITS-013807]

6. REFERENCES

1. Bharucha, J. J. Tonality and expectation. In R. Aiello & J. Sloboda (Eds.), Musical Perceptions (pp. 213-239). New York: Oxford University Press, 1994.
2. Boomsliter, P., & Creel, W. "The long pattern hypothesis in harmony and hearing." J. Music Theory, 5, 2-31, 1962.
3. Cariani, P. "Temporal codes, timing nets, and music perception." J. New Music Res., 30(2), 107-136, 2002.
4. Cariani, P. A., & Delgutte, B. "Neural correlates of the pitch of complex tones. I. Pitch and pitch salience." J. Neurophysiol., 76(3), 1698-1716, 1996a.
5. Cariani, P. A., & Delgutte, B. "Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase-invariance, pitch circularity, and the dominance region for pitch." J. Neurophysiol., 76(3), 1717-1734, 1996b.
6. Deutsch, D. The processing of pitch combinations. In D. Deutsch (Ed.), The Psychology of Music (pp. 349-411). San Diego: Academic Press, 1999.
7. DeWitt, L. A., & Crowder, R. G. "Tonal fusion of consonant musical intervals: The oomph in Stumpf." Perception and Psychophysics, 41(1), 73-84, 1987.
8. Geisler, C. D., & Greenberg, S. "A two-stage nonlinear cochlear model possesses automatic gain control." J Acoust Soc Am, 80(5), 1359-1363, 1986.
9. Green, B., & Butler, D. From acoustics to Tonpsychologie. In T. Christensen (Ed.), The Cambridge History of Western Music Theory (pp. 246-271). Cambridge: Cambridge University Press, 2002.
10. Hartmann, W. M. Signals, Sound, and Sensation. New York: Springer, 1998.
11. Helmholtz, H. v. On the Sensations of Tone as a Physiological Basis for the Theory of Music (1954 Reprint). New York: Dover, 1885.
12. Kameoka, A., & Kuriyagawa, M. "Consonance theory I. Consonance of dyads." J. Acoust. Soc. Am., 45, 1451-1459, 1969a.
13. Kameoka, A., & Kuriyagawa, M. "Consonance theory II. Consonance of complex tones and its calculation method." J. Acoust. Soc. Am., 45, 1460-1469, 1969b.
14. Kiang, N. Y. S., Watanabe, T., Thomas, E. C., & Clark, L. F. Discharge Patterns of Single Fibers in the Cat's Auditory Nerve. Cambridge: MIT Press, 1965.
15. Kim, D. O., & Molnar, C. E. "A population study of cochlear nerve fibers: comparison of spatial distributions of average-rate and phase-locking measures of responses to single tones." J. Neurophysiol., 42(1), 16-30, 1979.
16. Krumhansl, C. L. Cognitive Foundations of Musical Pitch. New York: Oxford University Press, 1990.

17. Leman, M. "An auditory model of the role of short-term memory in probe-tone ratings." Music Perception, 17(4), 481-510, 2000.

18. Leman, M., & Carreras, F. Schema and Gestalt: Testing the hypothesis of psychoneural isomorphism by computer simulation. In M. Leman (Ed.), Music, Gestalt, and Computing (pp. 144-165). Berlin: Springer, 1997.

19. Licklider, J. C. R. "A duplex theory of pitch perception." Experientia, VII(4), 128-134, 1951.

20. Lipps, T. Consonance and Dissonance in Music (1995 translation) (W. Thompson, Trans.). San Marino, CA: Everett Books, 1905.

21. Meddis, R., & Hewitt, M. J. "Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I. Pitch identification." J. Acoust. Soc. Am., 89(6), 2866-2882, 1991.

22. Meddis, R., & O'Mard, L. "A unitary model of pitch perception." J. Acoust. Soc. Am., 102(3), 1811-1820, 1997.

23. Parncutt, R. Harmony: A Psychoacoustical Approach. Berlin: Springer-Verlag, 1989.

24. Patterson, R. D. "Spiral detection of periodicity and the spiral form of musical scales." Psychology of Music, 14, 44-61, 1986.

25. Plomp, R., & Levelt, W. J. M. "Tonal consonance and critical bandwidth." J. Acoust. Soc. Am., 38, 548-560, 1965.

26. Pressnitzer, D., Patterson, R. D., & Krumboltz, K. "The lower limit of melodic pitch." J. Acoust. Soc. Am., 109(5), 2074-2084, 2001.

27. Rose, J. E. Neural correlates of some psychoacoustical experiences. In D. McFadden (Ed.), Neural Mechanisms of Behavior (pp. 1-33). New York: Springer Verlag, 1980.

28. Rose, J. E., Hind, J. E., Brugge, J. R., & Anderson, D. J. "Some effects of stimulus intensity on response of single auditory nerve fibers of the squirrel monkey." Journal of Neurophysiology, 34(4), 685-699, 1971.

29. Schneider, A. "Verschmelzung", tonal fusion, and consoance: Carl Stumpf revisited. In M. Leman (Ed.), Music, Gestalt, and Computing. Berlin: Springer, 1997.

30. Sethares, W. Tuning, Timbre, Spectrum, Scale. London: Springer, 1999.

31. Slaney, M., & Lyon, R. F. (1991). Apple Hearing Demo Reel [videotape and report]. Cupertino, CA: Apple Computer, Inc.

32. Snyder, B. Music and Memory. Cambridge: MIT Press, 2000.

33. Terhardt, E. "On the perception of periodic sound fluctuations (roughness)." Acustica, 30, 201-213, 1974a.

34. Terhardt, E. "Pitch, consonance, and harmony." J. Acoust. Soc. Am., 55(May), 1061-1069, 1974b.

35. Terhardt, E., Stoll, G., & Seewann, M. "Pitch of complex signals according to virtual-pitch theory: test, examples, and predictions." J. Acoust. Soc. Am., 71(3), 671-678, 1982.

36. Tramo, M. J., Cariani, P. A., Delgutte, B., & Braida, L. D. Neurobiology of harmony perception. In I. Peretz & R. J. Zatorre (Eds.), The Cognitive Neuroscience of Music (pp. 127-151 (Reprint of Tramo et al (2001), Annals NY Acad Sci, 2990:2092-2116)). New York: Oxford University Press, 2001.

Chapter 13
Toward a Theory of Information Processing in Auditory Cortex

Peter Cariani and Christophe Micheyl

13.1 Introduction

The primary goal of auditory science is to provide a full account of how humans and animals hear sounds of all kinds: the sounds of everyday life, environmental sounds, speech, and music. A comprehensive, neurally grounded theory of hearing is needed that explains precisely how we hear what we hear. This chapter discusses cortical function in the context of such a theory. The first half of the chapter (Sections 13.2 and 13.3) outlines the basic structure of hearing (what is to be explained) and the various aspects of neural information processing that are needed for adequate explanations of auditory function (the terms of the explanations). The second half (Section 13.4) lists some of the fundamental outstanding experimental and theoretical problems that need to be solved.

The goal of auditory theory is to understand how the auditory system works as an informational system. Once such a theory is finally formulated, such that a firm understanding of the codes, computations, and their neuronal substrates is achieved, then more effective therapeutic strategies for restoring auditory functions lost to disease can be devised, and artificial devices that expand the sound analysis capabilities of our own natural auditory systems can be designed and built.

What would a complete theory of audition entail? Such a theory would identify and account for the perceptual and cognitive *informational functions* that the auditory system carries out (Fig. 13.1), as well as the structure of subjective auditory

P. Cariani (✉)
Department of Otology and Laryngology, Harvard Medical School,
629 Watertown Street, Newton, MA 02460, USA
e-mail: cariani@mac.com

C. Micheyl
Auditory Perception and Cognition Laboratory, Department of Psychology,
University of Minnesota, N628 Elliot Hall, 75 East River Parkway,
Minneapolis, MN 55455, USA
e-mail: cmicheyl@umn.edu

D. Poeppel et al. (eds.), *The Human Auditory Cortex*, Springer Handbook of Auditory Research 43, DOI 10.1007/978-1-4614-2314-0_13,

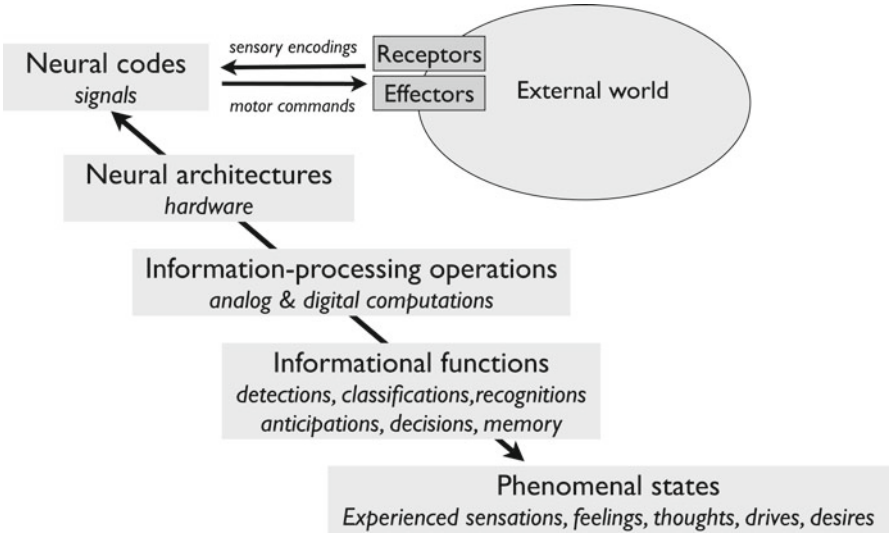

Fig. 13.1 Complementary aspects of neural information processing that are essential for understanding how the auditory system works. Auditory theory seeks to explain auditory functions and experiences in terms of neural codes, architectures, and information-processing operations. These aspects can be compared with Marr's (1982) division into computational, algorithmic, and implementational levels of description

experience (*phenomenal states*). Accounting for auditory functions in neural terms involves identification of neural codes, neuronal architectures, and neurocomputational operations. *Neural codes* are the basic signals of the system that support systematic internal representations of sound attributes. *Neuronal architectures* are the neuronal hardware substrates that implement neural signal processing. The rubric of architectures includes the composition and organization of neural elements, from molecular, cellular, and anatomical structures and functions to patterns of interneural connections. *Information-processing operations* on internal representations carry out transformations, analyses, and decisions that ultimately steer and switch behavior. These aspects of the auditory system tell us how the system must be organized so as to achieve its functions at any given time. However, an even fuller account of audition also includes a historical understanding of how the system came to be. This includes the ontogenetic, developmental processes that construct and modify the auditory system over the life of the organism and the phylogenetic, evolutionary processes that have shaped its structures and functions, from its first appearance in ancient animals to its present-day form.

As with many other structure–function relations in biology, each aspect is complementary to the rest. A given function, being a set of relations, does not uniquely determine the underlying structures and mechanisms that implement it. For example, auditory functions can potentially be carried out in different ways, using different neural codes, operations, and architectures that utilize different biophysical

mechanisms. Conversely, although underlying molecular, cellular, and network structures and biophysics constrain which codes, operations, and architectures are possible, complete knowledge of underlying neuronal substrates does not necessarily directly lead to better understanding of the codes, operations, and architectures that realize auditory functions.

Taken together, these complementary aspects of neural information processing yield a comprehensive, systematic account of how the system works: what functions it realizes, what kinds of internal signals and representations it uses, what processing operations are carried out, as well as the neural network architectures and properties of the neuronal system elements that are essential to its functioning. Such an account would constitute a full mechanistic and neurocomputational model of auditory structures, mechanisms, and functions.

Even so sweeping an account, however, would omit the phenomenal, subjective aspect of hearing that we as conscious subjects experience directly when we hear a sound. In the last two decades, there has been a resurgence of interest and investigation in the neuronal correlates of subjective experiences (Koch, 2004). The neural correlates of consciousness (NCCs) involve the neuronal requisites for having conscious awareness, while the neural correlates of the contents of consciousness (NCCCs) involve the specific neural basis for particular experiences.

Although most empirical investigation and theory-building in this emerging field of consciousness studies has involved the visual system, involving phenomena such as visual masking, diverted attention, blindsight, and visual neglect syndromes (Pollen, 2008), close auditory analogues of all these phenomena exist. Thus, there is no reason why the neural basis of auditory experience cannot furnish comparable insights into the neural basis for conscious awareness. What are the minimal, requisite neuronal conditions for a sound to be consciously experienced (auditory NCCs), and what patterns of neuronal activity correspond to which changes in auditory experience (auditory NCCCs)? These are fundamental questions that are also practically relevant for understanding the neurophenomenological basis of auditory hallucinations and tinnitus.

A full theory of audition thus should explain the relations between sounds, neuronal responses, auditory functions, and auditory experience (Fig. 13.2). Psychoacoustics and psychology address questions concerning relations between acoustic stimuli and auditory perception and cognition. Neurophysiology and computational neuroscience seek to characterize neuronal responses to sounds (system identification) and to identify correspondences between neuronal responses and auditory functions (neural coding) such that underlying information processing principles that are utilized by the auditory system can be identified (reverse engineering). Lastly, one can point to a future field of auditory neurophenomenology that would elucidate the structure of subjective auditory experience and formulate neurophenomenal bridge laws that predict the dimensional structure and contents of experience (NCCs and NCCCs) from neural activity patterns.

This chapter begins with the basic auditory functions that auditory theory seeks to explain (Section 13.2) followed by discussion of the nature of such an explanation in terms of the codes, computational operations, and architectures that might be

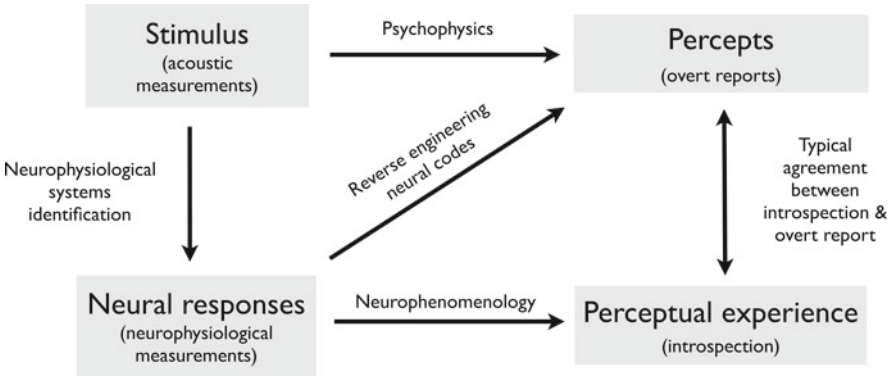

Fig. 13.2 Stimulus–response relations and neuropsychological states. Auditory psychophysics involves modeling of stimulus–percept relations. Neurophysiological systems identification involves prediction of the behavior of neural elements and networks as a function of the acoustic stimulus. The neural coding problem involves identifying, through reverse engineering, which aspects of neural response are causally related to informational (perceptual) functions. Neurophenomenology involves identifying which aspects of neural activity are necessary and/or sufficient to produce particular experiential states. Public and private measurements associated with acoustic, neural, behavioral, and phenomenal states are listed in parentheses

involved (Section 13.3). Succeeding sections then take up outstanding problems and fundamental questions about how the auditory system works and the role auditory cortex might play in that process. These questions involve identifying the neural codes operant at the cortical level and understanding the neurocomputational basis for invariances and invariant transformations of auditory percepts as well as their precision and robustness. This discussion seeks to present the fundamental scientific questions that auditory theory faces in as comprehensive and clear manner as possible, to provoke deeper thinking for more comprehensive theory-building that can guide experimental investigations. Unlike simple reports of empirical data, which usually stand on their own, these ideas are necessarily exploratory and speculative, and are meant to widen rather than narrow the realm of possible mechanisms for consideration.

13.2 What a Neurocomputational Theory of Auditory Cortex Seeks to Explain

13.2.1 General Auditory Functions

The primary goal of a theory of audition is to explain auditory function. This first entails a broad, ecological account of how the auditory system enhances the survival and reproductive fitness of the organism and its lineage as well as specific accounts

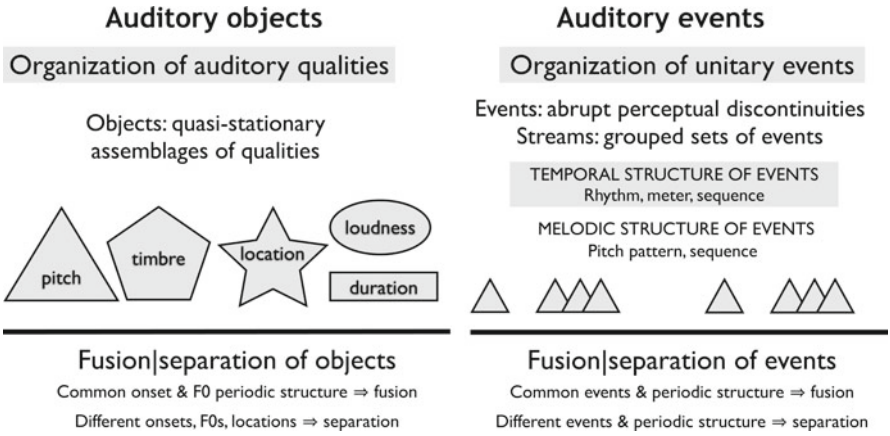

Fig. 13.3 Perceptual organization of auditory qualities and events. (Left) Grouping of auditory qualities in objects and major factors that govern the fusion or separation of objects. (Right) Grouping of events into streams and major factors that govern the fusion or separation of streams

of particular perceptual and cognitive functions that the auditory system realizes, such as detection, discrimination, object and stream formation and separation, analysis, and recognition of sounds. These specific functions include representation and discrimination of different physical aspects of sounds, such as periodicity, spectrum, intensity, duration, and their spatial relationships to listeners. They also include auditory functionalities related to inferences about the auditory scene (how many independent sound sources and the properties of the individual sounds) as well as learned past experiences of sound that have been retained in memory: familiarity, hedonic preference (euphonious vs. unpleasant attributes), category (e.g., phonetic class, or for absolute pitch possessors, pitch class), sound source identity (what speaker, musical instrument, animal, or natural process has produced a given sound?). These functions all provide critical information about environmental sound sources that facilitate their recognition. Auditory functions support more appropriate and effective responses to environmental events.

A comprehensive theory of auditory cortical function should explain how perception of different qualities of sound (pitch, timbre, loudness, duration, location) is subserved by neuronal cortical representations and mechanisms. How do human and animal listeners discriminate fine differences to differentiate sounds and generalize similarities to recognize categories of sound? How does auditory theory account for the dimensional structure of auditory perception—the nearly complete independence of different perceptual attributes? Such a theory should also explain why the auditory scene has the organization that it does, such that separate objects and events are distinguished and their respective attributes grouped together. The auditory scene at any given time contains a set of quasi-stable objects that have an organization in terms of their respective attributes (Fig. 13.3). Analogously, there is a larger organization of the auditory scene in time, in which longer temporal patterns of related events cohere into unified patterns (e.g., rhythmic and melodic sequences), with their separation and fusion into streams (stream segregations and fusions).

Basic auditory attributes are discussed first, followed by their organization in terms of object representations, and the temporal organization of auditory events and objects into streams.

13.2.2 Organization of the Auditory Scene

Perception has a strong dimensional structure, both at the level of different sensory modalities and kinds of distinctions within each modality (Boring, 1942). In humans, major sensory modalities include audition, vision, olfaction, gustation, balance and orientation, proprioception, pain, thermoreception, flutter-vibration, pressure, as well as a host of various interoreceptors. In turn, audition has a strong dimensional structure that is reflected in basic attributes of sound perception: loudness, pitch, timbre, duration, and apparent location. There is also a higher-level organization of these primitive auditory qualities in which some sets of attributes are grouped together in objects (Fig. 13.3). At any present moment, auditory experience consists of a temporal succession of perceived objects and events. Auditory scene analysis involves accounting for how the auditory system organizes incoming sound patterns into auditory objects and events, each with its own set of associated perceptual attributes (Handel, 1989; Bregman, 1990). The general principles that underlie auditory perceptual organization were originally investigated by Gestaltist psychologists (e.g., Köhler, 1947), and later came to be popularly known as the "cocktail party problem" (Cherry, 1966). Such organizing principles play a particularly important role in structuring musical experiences and expectations (Handel, 1989; Bregman, 1990; Snyder, 2000).

An auditory *object* is a representational structure that is a collection of stable, quasi-invariant perceptual properties (attributes, features) that persist together over time as a unitary entity. In contrast, an auditory *event* is a representation of a salient change in the perceived auditory scene from one moment to the next that also has a set of properties (attributes, features, contrasts) associated with it. Auditory objects and events are mental constructs that may or may not correspond directly to particular, identifiable physical objects, sound sources, or sonic events in the external world. Such constructs are produced by internal representational structures that are implemented through organized patterns of neuronal activity.

A prime example of an auditory object is the perception of a single note played by a musical instrument (the temporal stability of the note being emphasized), whereas an example of an auditory event could involve perception of different, temporal aspects of the same note (the temporal succession of the note-object's appearance and disappearance being emphasized). One can also consider the onset and offset of the note as separate auditory events in and of themselves. Auditory events are mental constructs typically produced by temporal acoustic contrasts that distinguish subsequent from preceding sound patterns. The perception of a discrete event is itself the result of a temporal auditory grouping process. In most music, the onset of each note is an individual auditory event. Event boundaries in speech can be more fluid, where individual phonetic elements or rapid sequences of "chunked" elements can constitute separate, unified events.

13.2.3 Basic Auditory Qualities

Perceptual auditory qualities associated with each object or event can be grouped under more general categories of pitch, timbre, location, loudness, and duration (Fig. 13.3, left). In turn, each category can include one or more related perceptual dimensions. Some attributes, such as loudness and duration, are one dimensional, such that their qualities can be ordered in a monotonic, linear order. Location here includes several spatial qualities of sounds that include the apparent direction, extent, and range of the sound image in auditory space. Other qualities, such as pitch and timbre, are multidimensional, having several related, but distinct aspects that may be related to different underlying neural representations.

This high level division of auditory qualities is most obvious in musical contexts. Operationally, pitch is that auditory quality that covaries with the frequency of pure tones. Instruments are generally recognizable by their distinctive timbres, irrespective of the pitch, loudness, and duration of the notes played and of their position in auditory space. Instruments that produce extremely different timbres can play notes with the same recognizable pitches, and the pitches and timbres that they produce are generally highly invariant with respect to sound intensity. Although extremely short sound durations can alter perception of pitch, timbre, and loudness, these qualities are highly invariant for longer durations (>50 ms). Whereas tonal music typically involves melodic sequences of auditory events (notes) having varying pitches with relatively fixed timbres, speech communication systems typically use sequences of distinctive auditory events (phonetic elements) that have different timbres.

13.2.3.1 Loudness, Duration and Spatial Attributes

Each auditory quality has primary acoustical correlates. Loudness covaries monotonically with intensity, perceived duration with duration. Directionality in the horizontal, azimuthal plane depends on interaural time-of-arrival and sound pressure level differences at the two ears. Perceived elevation depends on high-frequency spectral notches that are characteristic of a listener's pinnae. Many physical correlates of attributes associated with spatial hearing, such as apparent distance, source width/extent, and enclosure size, are joint properties of sound sources and their reflections (Ando & Cariani, 2009).

13.2.3.2 Pitch

Pitch is somewhat more complex in structure. Operationally, the pitch of a sound is the perceptual quality that covaries with the frequency of pure tones, that is, it is defined as the frequency of a pure tone to which the perceived pitch quality of given stimulus is matched. Pitch is related to the dominant periodicity of sounds, the repetition rate of a sound pattern. For pure tones this repetition rate is simply the tone's frequency; for complex tones this repetition rate is the fundamental frequency (de Cheveigné, 2005).

Arguably there are two closely related percepts that are associated with pitch, here called "spectral pitch" and "periodicity pitch." Differences in some of their properties suggest that these percepts may be subserved by qualitatively different neural representations. Pitch height is related to the perception of the absolute "lowness or highness" of a spectral pitch, and is monotonically related to the absolute frequencies of dominant spectral components present in a sound.

Periodicity pitch is a quality related to the dominant periodicity of a sound, its fundamental frequency, irrespective of its spectral composition. It is often also called "virtual pitch," "the low pitch of complex tones," "F0 pitch," "the pitch of the (missing) fundamental," or "musical pitch," depending on context and theoretical orientation.

Low-frequency pure tones evoke both spectral and periodicity pitch. Harmonic stimuli, such as AM tones, can be constructed in which pitch height (which varies with carrier frequency) and periodicity (which varies with modulation frequency) can be independently altered. Both spectral and periodicity pitches support local judgments of whether one pitch is lower or higher than another.

Pitch height and periodicity pitch appear to reflect different, independent acoustic properties, that is, absolute frequencies vs. dominant periodicities, that are likely to be mediated by different neuronal representations. At the level of the auditory nerve spectral pitch appears to be based on cochlear place of maximal excitation, whereas periodicity pitch appears to be based on spike timing information, most likely in the form of population-wide distributions of interspike interval (Cariani & Delgutte, 1996; Cariani, 1999). The cochlear place representation runs the entire frequency range of hearing (~50–20,000 Hz), whereas useable spike timing information in the auditory nerve extends only up to the limit of significant phase locking, around 4–5 kHz. At the cortical level, changes in pitch height and periodicity-pitch chroma produce neural activity in different local areas of human auditory cortex (Warren et al., 2003).

Musical tonality appears to be related to periodicity pitch and the temporal representation. Most listeners readily recognize octave similarities, musical intervals, melodies and their transpositions, as well as mistuned, and "off-key" or mistuned "sour" notes for periodicities below 4–5 kHz, that is, within the existence region for periodicity pitch. These distinctions all involve relative pitch, which involve recognitions of particular ratios of periodicities (see McDermott & Oxenham, 2008, for discussion of cortical mechanisms and Section 13.4.4 for discussion of relative pitch and melodic transposition). Although very crude melodic distinctions based on pitch height contours can be made for sequences of pure tones above this frequency limit, in the absence of periodicity pitch, recognitions of melodies and musical intervals in this register are severely degraded. Whereas periodicity pitch is highly invariant with respect to sound intensity, spectral pitch perception of high-frequency pure tones is notoriously level dependent, consistent with the notion that the two types of pitch percepts depend, respectively, on temporal and cochlear-place neural representations. Critical questions for auditory neurophysiology involve the nature of neural coding transformations that give rise to cortical representations for spectral and periodicity pitch.

Pitch strength or salience is an intensive dimension of pitch that is related to the apparent strength of a pitch. Pitch strength is analogous to color saturation in vision. Bandwidth is the primary acoustical determinant of spectral pitch strength, whereas waveform regularity/harmonicity and harmonic number are primary determinants of the strengths of periodicity pitches. These factors been used to parametrically vary periodicity pitch strength, such that one can identify cortical territories whose neural populations respond differentially to stimuli that evoke periodicity pitches (Patterson et al., 2002; Warren et al., 2003; Penagos et al., 2004).

13.2.3.3 Timbre

Timbre is perhaps the most complex perceptual category, with multiple dimensions and acoustical correlates. Timbre is most clearly illustrated in musical contexts, where it includes those qualities of sound that distinguish the same musical note (i.e., of the same pitch, duration, and loudness) played by different instruments at the same location. Some aspects of timbre depend on the gross power spectrum of stationary sounds (e.g., spectral center of gravity, spectral tilt/brightness, formant structure, and bandwidth), whereas other aspects depend on rapid modulations and fluxes in amplitude, frequency, and phase of nonstationary sounds (Handel, 1989; McAdams & Giordano, 2009).

In musical contexts, instrument resonances determine aspects of timbre related to the power spectrum (tone color, brightness), while onset and offset dynamics (attack, sustain, decay, tremolo), frequency fluxes (temporal successions of harmonics, vibrato, noise components), and phase dynamics (chorus, flanger and phaser effects) determine those aspects of timbre that are related to rapid changes in sound.

By this expansive definition of timbre, most phonetic distinctions in speech are categorical timbral distinctions. Phonetic distinctions in atonal languages either involve the gross power spectra of stationary sounds (vowels) or amplitude and frequency transients (consonants). In tonal languages, pitch levels and contours are also used alongside timbral differences to distinguish phonetic elements.

Whereas timbral space associated with stationary spectral distinctions is relatively well understood, the structure of timbral space associated with transient changes in amplitude, frequency and phase has not been systematically characterized. As with pitch, the multiplicity of timbral qualities may be associated with different aspects of neural patterns of response.

13.2.4 Formation of Auditory Objects and Events

Auditory perception consists of more than just elemental perceptual qualities—there is an organization of perceptual attributes within the "auditory scene" in which neural representations associated with particular sound components are grouped together to form representations of unitary auditory objects or events (Handel, 1989;

Bregman, 1990). The auditory scene at any particular moment can contain multiple auditory objects, with each object having its own set of associated perceptual and cognitive attributes (loudness, duration, location, pitch, timbre plus higher-order cognitive attributes) that group together (Griffiths, Micheyl, and Overath, Chapter 8).

Common harmonic structure and common onset time are the two strongest factors that respectively cause sets of frequency components to fuse into unified auditory objects and events (Fig. 13.3, bottom left). In comparison, the sound parameters associated with location, duration, and loudness produce a much weaker basis for grouping. Common harmonic structure means that the sound patterns and internal early temporal representations of groups of harmonically related frequency components form repeating, periodic patterns. Common onset of components means that the simultaneous sound patterns, whether harmonic or inharmonic, will produce the same frequency–time pattern of spike timings and firing rates. This generates a characteristic timbre for the auditory event. Thus when the same components recur at a later time, the common onset grouping has a similar timbral and pitch representation. When frequency components have neither common harmonic structure nor common onset, the mixture is inharmonic and the representations of their sound patterns do not fuse, such that multiple pitches may be heard. Thus, one can hear out the notes of several musical instruments playing at the same time, provided that the notes are not harmonically related and their onsets are not strictly synchronized. In such cases when the notes form separate objects and events, their respective qualities (pitch, timbre, loudness, duration, and location), can be heard, such that individual instruments can be identified by the timbres of the individual notes. On the other hand, some sounds (multiple broadband noises or harmonic complexes with the same fundamental frequency) fuse together into one auditory object. To the extent that different sounds have separate internal representations, they have separate sets of perceptual attributes; to the extent that they are fused together into a single representational object, their attributes become blended together. Thus the auditory scene is perceived as containing multiple auditory objects each with its own set of perceptual attributes.

13.2.5 Grouping of Events into Streams

Related auditory events in turn are grouped into distinct associated patterns called streams (Handel, 1989; Bregman, 1990). In a manner analogous to the formation and separation of objects, neural mechanisms group recurring patterns of events and sequences of similar events (e.g., common pitch, timbre, duration, location) into streams. Each stream has attributes related to the relations between the events of the stream (Fig. 13.3, bottom right). Melody is a temporal pattern of pitch changes that coheres together as a unified, recognizable sequence. If the notes are too short (<100 ms) they blend together; if they are too long (>2–3 s), only one note is retained in echoic memory and the pitch pattern is not perceived. Similarly, rhythm is a temporal pattern of the onsets and offsets of related events that are grouped together into

a stream. Like melody, if the events are too closely spaced temporally, they fuse together, and the patterns of intervening time intervals are lost. If the events are too far apart, only individual events are perceived and no pattern is established. Although longer patterns of events do not cohere into palpable rhythms, repeating sequences of events on a wide range of timescales can be recognized after several cycles have been heard.

On all timescales, repeating patterns of sound and their evoked auditory events build up strong representational expectancies of their continuation (Snyder, 2000; Winkler et al., 2009). The effect is created even with arbitrary and highly artificial repeating sound patterns. Whatever the constancies or changes, repeating sequences build up strong auditory expectations of their continuation. When sequences deviate from their previous repeated patterns, a perceptually salient expectancy violation is produced.

A great deal of music relies on the creation of musical tonal and rhythmic expectancies and their violations (Handel, 1989; Bigand, 1993; Zatorre and Zarate, Chapter 10). The push–pull of violation-induced tension and the confirmation of expectancies lies at the basis of "emotion and meaning" in complex, program music that is meant for intent listening (Huron, 2006). Such expectancy violations create distinct neural signatures (Trainor & Zatorre, 2009), such as the mismatch negativity (MMN) (Näätänen et al., 2007) and other responses (Chait et al., 2007). Thus far, it appears that any perceptible deviation from an expected pattern, such as changes in loudness, pitch, timbre, duration, or event timing, creates a delayed mismatch negativity response. Such negativities are also seen for cognitive expectancies: linguistic syntactic and semantic violations as well as deviations from musical expectancies (Patel, 2008). The various negativity responses differ in their latencies relative to the time of the violation, a reflection of the neuronal populations involved and their interconnections with auditory cortical populations. These response similarities that involve different types of perceptual and cognitive attributes suggest the existence of general cortical mechanisms for the buildup of temporal pattern expectancies and their comparison with incoming temporal sequences.

13.2.6 Cognitive Dimensions

In addition to basic perceptual auditory attributes, there are also additional, cognitive aspects of sounds that are related to internal representations and past experiences. Representations associated with these cognitive dimensions may have their own grouping and comparison mechanisms that involve both auditory and nonauditory centers. These dimensions include:

- *Categorical perception.* Sound objects and events can be recognized as categorical tokens in learned symbol systems (e.g., phonetic elements in speech, pitch classes for possessors of absolute pitch).
- *Mnemonics.* A sound can be perceived as familiar or unfamiliar, depending on whether it interacts strongly with specific short-, intermediate-, or long-term

memories. Such memories also encode learned statistical dependencies between sounds.

- *Semantics.* Sounds can acquire meaning from previous experience. Sounds can be associatively linked with perceived objects and events, such that subsequent presentation and recognition engages association-related neural anticipatory-prediction mechanisms.
- *Pragmatics.* Sounds are also experienced in the context of the internal goals and drives, such that sounds can acquire relevance for goal attainment and drive reduction (e.g., a dinner bell).
- *Hedonics.* A sound can be experienced as pleasant, neutral, or unpleasant. The hedonic valence of a sound can be related to purely sensory factors (e.g., grating or very high pitched sounds) or learned associations (e.g., the sound of a bell that precedes a shock).
- *Affective dimensions.* Beyond simple pleasantness or unpleasantness, sounds and sound sequences can induce particular emotional states.

13.3 Toward a Neurocomputational Theory of Auditory Cortex: Auditory Codes, Representations, Operations, and Processing Architectures

The previous section outlined major features of auditory perception and cognition ("what we hear") that a full theory of information processing by auditory cortex should ultimately explain. Explanation of "how we hear what we hear" is framed here in terms of neural representations, operations, and processing architectures.

Identification of the neural correlates of the basic representational dimensions and organizations of auditory perception and cognition is a critical step in development of a working theory of auditory cortex. The Gestaltist concept of neuropsychological isomorphism (Köhler, 1947) is a useful working hypothesis, that is, that every dimension of auditory perceptual function and experience reflects the dimensional structure of underlying neural representations and information-processing operations on which it depends. If so, then not only do neural representations and computations explain the structure of perception and of experience, but these structures also provide strong clues as to the nature of the underlying neuronal processes.

13.3.1 Neural Codes

Neural codes involve those aspects of neuronal activity that play functional and informational roles in the nervous system, that is, they are specific patterns of activity that switch the internal states, and ultimately the overt behavior of the system (Rieke et al., 1997). Many different kinds of neural pulse codes are possible (Cariani, 1995), and whole catalogs of possible neural codes and evidence for them have been

discussed and collated in the past (Perkell & Bullock, 1968). Neural coding of sensory information can be based on discharge rates, interspike interval patterns, latency patterns, interneural discharge synchronies and correlations, temporal spike-burst structure, or still more elaborate cross-neuron volley patterns. In addition, sensory coding can be based on the mass statistics of many independent neural responses (population codes) or on the joint properties of particular combinations of responses (ensemble codes).

Amidst the many ways that neural spike trains can convey sensory information are fundamentally two basic ideas: "coding-by-channel" and "coding-by-time" (Fig. 13.4, top). Channel-based codes depend on the activation of specific neural channels or of configurations of channels. Temporal codes, on the other hand, depend on the relative timings of neural discharges rather than on which particular neural channels respond how much. Temporal codes can be based on particular patterns of spikes within spike trains (temporal-pattern codes) or on the relative times-of-arrival of spikes (time-of-arrival codes).

13.3.1.1 Channel-Based Coding Schemes

Many different channel-based coding schemes are possible. Such schemes can range from simple, unidimensional representations to low-dimensional sensory maps to higher dimensional feature detectors. In simple "doorbell" or "labeled line" systems, activation (or suppression) of a given neuron signals the presence or absence of one particular property. In more multipurpose schemes, neurons are differentially tuned to particular stimulus properties, such as frequency, periodicity, intensity, duration, or external location. Profiles of average discharge rates across a population of such tuned elements then convey multidimensional information about a stimulus. When spatially organized in a systematic manner by their tunings, these elements form sensory maps, in which spatial patterns of channel activation can then represent arbitrary combinations of those stimulus properties. In lieu of coherent spatial order, tuned units can potentially convey their respective channel identities through specific connections to other neurons beyond their immediate neighborhood. More complex constellations of properties can be represented via more complex concatenations of tunings to form highly specific "feature detectors." In the absence of coherent tunings, combinations of idiosyncratic response properties can potentially support "across-neuron pattern codes" of the sort that have been proposed for the olfactory system.

Nevertheless, idiosyncratic across-neuron patterns and associative learning mechanisms present fundamental difficulties in explaining common strong perceptual equivalence classes that are shared by most humans and are largely independent of an individual's particular history. Although these various functional organizations, from labeled lines to feature detectors to across-neuron patterns, encompass widely diverse modes of neural representation, all draw on the same basic strategy of coding-by-channel. In channel-coding schemes, it is usually further assumed that distinctions between alternative signal states are encoded by different average discharge rates. The combination of channel- and rate-based coding has remained by

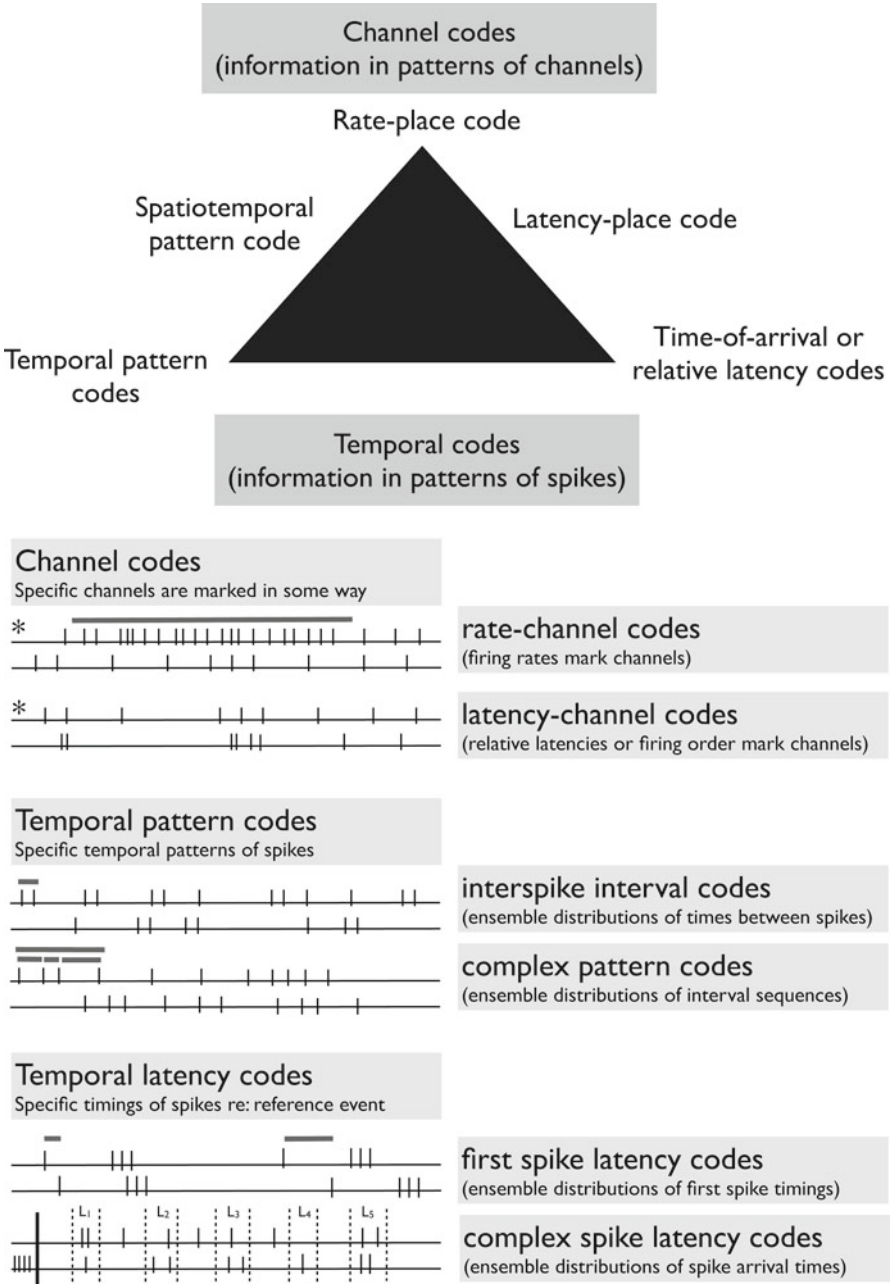

Fig. 13.4 Basic types of neural pulse codes. (Top) Division of codes into channel codes, temporal pattern codes, and relative latency codes. The three types are complementary, such that combination codes can be envisioned (e.g., marking of channels by particular spike patterns, spike latencies, or firing order rather than average rate). (Bottom) Schematic illustration of different code types. Channel codes convey information via patterns of marked channels (*), whereas temporal codes

far the dominant neural coding assumption throughout the history of neurophysiology (Boring, 1942), and, consequently, forms the basis for most of our existing neural-network models.

Within channel-coding schemes, aspects of the neural response other than rate, such as relative latency or temporal pattern, can also play the role of encoding alternative signal states Combination latency-place and spatiotemporal codes are shown in Fig. 13.4. In a simple latency-channel code, channels producing spikes at shorter latencies relative to the onset of a stimulus indicate stronger activation of tuned elements. Patterns of relative first-spike response latencies can encode stimulus intensity, location, or other qualities (Eggermont, 1990; Brugge et al., 1996; Heil, 1997). Temporal, channel-sequence codes have also been proposed in which the temporal order of neural response channels conveys information about a stimulus (Van Rullen & Thorpe, 2001).

Common-response latency, in the form of interchannel synchrony, has been proposed as a strategy for grouping channels to form discrete, separate objects (Singer, 2003). In this scheme, rate patterns across simultaneously activated channels encode object qualities, whereas interchannel synchronies (joint properties of response latencies) create perceptual organization, which channels combine to encode which objects. The concurrent use of multiple coding vehicles, channel, rate, and common time-of-arrival permits time-division multiplexing of multiple objects. Still, other kinds of asynchronous multiplexing schemes are possible if other coding variables, such as complex temporal patterns and temporal pattern coherences, are used (Emmers, 1981; Cariani, 2004; Panzeri et al., 2009).

13.3.1.2 Temporal Coding Schemes

Characteristic temporal discharge patterns can also convey information about stimulus qualities. Neural codes that rely predominantly on the timings of neural discharges have been found in a variety of sensory systems (reviews: Cariani, 1995, 1999, 2001b, 2004). Conceptually, these temporal codes can be divided into time-of-arrival and temporal-pattern codes (Fig. 13.4).

Fig. 13.4 (continued) convey information via patterns of spikes (bars). In rate-place codes, across-neuron firing rate patterns of tuned elements convey information (e.g., coding of stimulus power spectra via rate profiles of frequency-tuned neurons). Temporal pattern codes use temporal patterns among spikes, such as distributions or sequences of interspike intervals, to convey information (e.g., coding of periodicity pitch via all-order interspike interval distributions). Time-of-arrival or relative latency codes use relative arrival times of spikes to convey information (e.g., coding of azimuth via spike timing disparities between left and right auditory pathways). They also can use distributions of timings of spikes with different latencies following some common initial event (vertical bar) to encode the selective activations of various neuronal assemblies that have different response latencies (L_1–L_4)

Time-of-arrival codes use the relative times of arrival of spikes in different channels to convey information about the stimulus. Examples of time-of-arrival codes are found in many sensory systems that utilize the differential times of arrival of stimuli at different receptor surfaces to infer the location of external objects (von Békésy, 1967; Carr, 1993). Strong examples are auditory localizations that rely on the time-of-arrival differences of acoustic signals at the two ears, echolocation range findings that rely on time-of-arrival differences between emitted calls and their echoes, and electroceptive localizations that use the phase differences of internally generated weak electric fields at different locations of the body to infer the presence of external phase distortions caused by nearby objects.

Temporal pattern codes, such as interspike interval codes, use temporal patterns between spikes to convey sensory information. In a temporal pattern code, the internal patterns of spike arrivals bear stimulus-related information. The simplest temporal pattern codes are interspike interval codes, in which stimulus periodicities are represented using the times between spike arrivals. More complex temporal pattern codes use higher-order time patterns consisting of interval sequences (Emmers, 1981; Abeles et al., 1993; Villa, 2000). Like time-of-arrival codes, interval and interval-sequence codes could be called correlational codes because they rely on temporal correlations between individual spike-arrival events. Temporal pattern codes should be contrasted with conceptions of temporal coding that rely on temporal variations in average discharge rate or discharge probability. These temporal-rate codes count spikes across stimulus presentations as a function of time and then perform a coarse temporal analysis on changes in spike rates.

Both time-of-arrival and temporal-pattern codes for conveying sensory information depend on spike timing patterns that are characteristic of a given stimulus attribute. The stimulus-related temporal discharge patterns on which temporal-pattern codes depend can arise in two ways: through direct stimulus locking and through stimulus-triggered, intrinsic-time courses of response (i.e., characteristic impulse response forms of receptors, sensory peripheries, and central neuronal assemblies). Some temporal codes permit signal multiplexing (Cariani, 1995, 2001b, 2004; Panzeri et al., 2009), such that different types of information can be transmitted concurrently over the same axonal transmission lines.

13.3.2 Neural Representations

Neural representations are patterns of neural activity that provide systematic means of encoding a set of informational distinctions. From psychoacoustic studies and our own direct experiences as listeners, there appear to be informational structures that provide for systematic representation of parameters that are associated with the basic auditory qualities of loudness, duration, location, pitch, and spatial hearing (directionality, apparent size). The concept of the neural code emphasizes the specific aspect of neuronal activity (e.g., firing rates, spike correlation patterns, relative latencies), whereas the concept of a neural representation emphasizes the systematic

nature of the distinctions being conveyed. For example, the neural representation of sound direction relies on a spike timing code based on relative latency at the level of the auditory brain stem, but the cortical representation might be based on patterns of spike rates or first spike latencies across direction-tuned neural populations (Brugge et al., 1996). The existence of coherent internal representations for different aspects of sounds is inferred from the systematic nature of perceptual judgments, and the relative independence of judgments related to these different aspects. For example, two very different sounds, in terms of spectrum and periodicity, can nevertheless be compared in terms of their loudness, apparent location, and duration. Although there can weak interactions between these dimensions, for most low-frequency hearing involving speech and music, the dimensions are remarkably independent. The different, independent representations are the presumed basis for the dimensional structure of the percept space.

Although many different kinds of neurophysiological and neuroanatomical findings suggest that particular kinds of representations are likely to be utilized by the auditory system, the only way that one can reliably test whether a given pattern of neural activity serves as the informational vehicle for some perceptual or cognitive function is to attempt to predict the percepts experienced directly from the neural data. If the putative neural codes and representations can be used to successfully predict specific patterns of perception and cognition, given additional neuroanatomical and neurocomputational constraints, then this constitutes strong evidence that the system itself is utilizing information in this particular form.

13.3.3 High-Level Information-Processing Operations

To realize perceptual functions, the auditory system carries out operations on the encoded information in representations. These operations transform patterns of neural activity that bear information (codes and representations) into decisions that then select subsequent action. Sets of information processing operations thus realize perceptual and cognitive functions such as

- detections (e.g., judgment of presence or absence of a sound pattern, “feature detections” of all sorts)
- comparisons or discriminations (e.g., estimating the similarity of two sounds, how they are alike or different, distinguishing two different pitches or note durations)
- classification and recognition (e.g., classification of phonetic elements, recognition of a familiar word or voice)
- anticipatory predictions (e.g., producing expectations of what sounds will occur next based on what has occurred before, on multiple timescales)
- object/stream formation and separation (e.g., hearing out individual voices from a mixture)
- tracking of objects/streams (e.g., following the apparent direction of a moving sound source)

- attention (e.g., focusing on particular objects, streams or aspects of sounds, enhancement of some representations and suppression of others)
- "evaluation," (e.g., assessing the relevance of a given sound to survival, reproduction, or more specific, current goals)

These information processing operations are carried out by neurocomputational mechanisms that are discussed in the next section. Many mechanisms for comparing sound patterns and attributes involve memory mechanisms on different timescales.

13.3.4 Low-Level Neurocomputations

Neurocomputations are processes on the level of individual neurons that carry out the most basic signal processing operations. Examples of basic neurocomputations include:

- thresholding operations (e.g., spike generation)
- spike addition and temporal integration (i.e., spatial and temporal summation of excitatory inputs)
- subtraction (excitatory vs. inhibitory inputs)
- coincidence detection (spike multiplication)
- anticoincidence (spike disjunction via coincidence detection with excitatory and inhibitory inputs)
- time delay (synaptic delay, conduction delay, inhibitory rebound)
- membrane threshold accommodation (high-pass filtering, onset detection)
- spike-pattern generation (e.g., bursting patterns)
- axonal spike-train filtering (via activity-dependent conduction blocks)
- synaptic functional modification (e.g., spike-timing-dependent plasticity)

Out of these and other biophysical processes, signal processing elements such as leaky integrators, coincidence detectors, and onset detectors can be constructed. From combinations of these basic sets of computational primitives, even more complex operations can be realized.

A useful, concrete example can be found in the basic neurocomputational operations in the auditory brain stem that subserve auditory localization from acoustic interaural time-of-arrival differences. Here the neural code is a spike timing (relative latency) code that is a consequence both of differences in sound arrival time at the two ears and of phase-locking of low frequency components. A binaural cross-correlation operation is carried out in the auditory brain stem using axonal delay lines, precisely timed inhibition, and coincidence detection in bilaterally symmetric, bipolar neurons. Thus, utilizing biophysical mechanisms and dedicated, specialized neuroanatomical structures, this neural architecture implements a binaural cross correlation operation that supports systematic representation of the horizontal plane in auditory space.

13.3.5 Neural Architectures

A neural architecture is an organization of neural elements, including their interconnections and element response properties, which provide the anatomical and physiological substrate needed to implement basic neurocomputations. In turn, these neurocomputations realize information-processing functions, and ultimately perceptual and cognitive functions. Each kind of neural coding scheme requires a compatible neural architecture (what circuit organizations and element properties are available) for its implementation. Thus the question of the nature of neural codes that are operant in auditory cortex is intimately related to the question of the nature of the neurocomputations that are realized by the neural populations in auditory cortex.

What kind of neural information processing architecture is auditory cortex? There are several broad alternatives: rate-place connectionist architectures with or without spatial maps, synchronized or oscillatory connectionist architectures, time delay neural networks, or timing nets. The relative uniformity of cortical organization suggests that one basic architectural type handles all different kinds of incoming information, albeit with plastic adjustments that depend on the correlation structure of the inputs. Within the constraints given by coarse genetic specifications, the stimulus organizes the fine structure of the tissue. In all of these neural network types, network functions can be adaptively modified by changing synaptic efficacies and other biophysical parameters. Given this plasticity, it is almost certain that auditory cortex configures itself in different ways according to the different kinds of information that are determined by connections to other parts of the system (sensory surfaces, subcortical afferent pathways, descending pathways, other cortical and subcortical populations). It is conceivable that auditory cortex can support several, perhaps all, of these alternative processing organizations.

13.3.5.1 Connectionist Architectures

Rate-place connectionist architectures are neural networks in which all processing involves analysis of firing rate profiles among neural channels (“units” or “nodes”). The cerebral cortex is commonly regarded as a large recurrent connectionist Hopfield network whose informational states are N-dimensional vectors that represent the firing rates of its neural elements (e.g., Trappenberg, 2002). Because firing rates are scalar quantities, all informational distinctions must be made via different combinations of neural channel activations. Thus, the most basic assumption of connectionist systems is channel-coding (which channels are activated how much).

In a connectionist network, the signals emitted by each channel are “labeled” by virtue of their specific intranetwork connectivities that in turn determine their simple and complex tuning properties. In auditory cortex, neurons are thought to convey different kinds of information depending on their various tuning properties, such as selectivity for

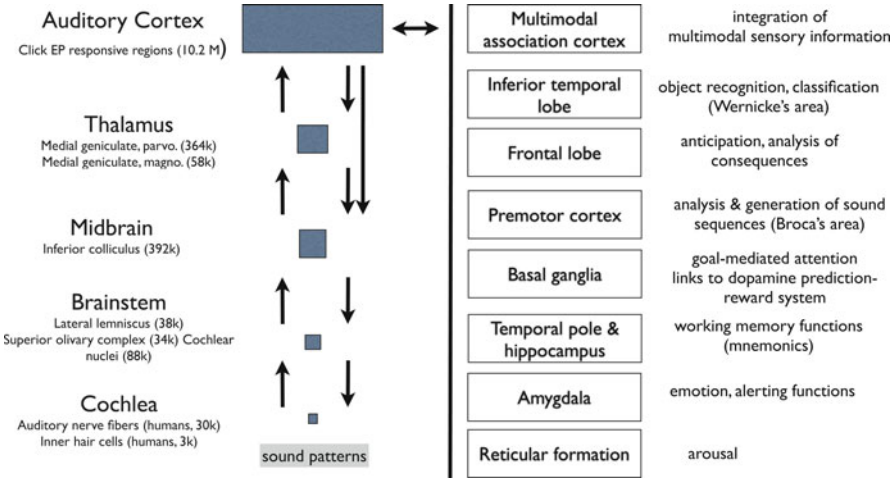

Fig. 13.5 Functional connections of auditory cortex to auditory pathway and the rest of the brain. (Left) Major levels in ascending and descending auditory pathways. Except where noted, numbers associated with auditory structures indicate numbers of neurons in the squirrel monkey auditory system (Chow, 1951). (Right) Major projections between auditory cortex and other brain structures, along with their basic functions. Connections between these structures and subcortical auditory pathways have been omitted

frequency, periodicity, sound location, bandwidth, intensity, or their modulations. The various tuning properties of a given neuron in turn depend on how it is connected to other neurons in the auditory system and other parts of the brain (Fig. 13.5).

In this vein, a number of studies have used linear system-identification techniques to characterize the time-frequency tuning properties of cortical neurons in terms of spectrotemporal receptive fields (STRFs) (e.g., Miller et al., 2002). Ideally, one should be able to use STRFs to predict the running firing rates of characterized neurons, ensembles, and populations to novel, complex stimuli. In practice, many neural elements behave unpredictably, with nonlinear responses that can change dynamically depending on recent stimulus history (Fritz et al., 2003).

There has been an ongoing debate about the nature of cortical processing elements, whether they are rate-integrators that are more compatible with connectionist schemes, or coincidence detectors operating on some kind of temporal code. If the functioning of auditory cortex does in fact depend on channel-coding schemes that use elements with relatively fixed receptive fields, be they dense or sparse, one might a priori expect the elements to have more reliable behavior (less discharge rate variance). On the other hand, all optimality arguments about neural codes and architectures are very risky to invoke at this point, before a reasonably firm grasp of how the system works has been attained. Given multitudes of neural elements, pooling of firing rate information via statistical population codes could potentially reconcile this apparent incongruity (see Section 13.4.2), but concrete mechanisms for pooling this information have yet to be identified. So far, no definitive answer has emerged.

Perhaps even more challenging for connectionist networks are problems of simultaneously representing and analyzing multiple auditory objects and event streams. More flexible kinds of networks are clearly needed to handle the combinatorics of multiple objects and their associated attributes. Temporal correlations between spike patterns (von der Malsberg, 1994) and emergent synchronies between spikes (Singer, 2003) could serve to bind together various feature detector channels that would group together corresponding attributes of auditory and visual objects. Synchrony-based grouping mechanisms have been the focus of much neurophysiological study in the visual cortex, albeit with equivocal correspondences with perception. Along similar lines, synchronized oscillations of neural firing have been proposed as auditory mechanisms for grouping channel-coded features and separating multiple sounds (Wang, 2002).

13.3.5.2 Oscillatory Networks

Neuronal oscillations have long been considered as potential mechanisms for informational integration (McCulloch, 1951; Greene, 1962). Stimulus-driven, stimulus-triggered, and endogenous, intrinsic oscillations are widespread in the brain (Buzsáki, 2006). Stimulus-driven oscillations follow the time structure of the stimulus, whereas stimulus-triggered oscillations, although evoked by an external stimulus, have their own intrinsic time courses that can also convey information about the stimulus (Bullock, 1992; Thatcher & John, 1977). Emergent, stimulus-triggered oscillations have been observed in olfactory systems and in the hippocampus, where spike latencies relative to oscillatory field potentials plausibly encode respectively, odor qualities (Laurent, 2006) and positional information relevant for navigation. These kinds of phase- or latency-based codes can either support marking of specific subsets of channels or ensemble-wide readouts of complex temporal patterns of response latency (Fig. 13.4). General purpose oscillatory-phase-latency codes for encoding signals and rhythmic-mode processing mechanisms for integrating multimodal information have been proposed (Schroeder & Lakatos, 2009).

Despite widespread evidence for oscillatory coupling of many neuronal populations it is not yet clear whether the various gamma, theta, and alpha oscillations that are seen in cortical populations play obligatory or specific informational roles as either temporal frameworks for phase-precession codes or channel-grouping mechanisms. Instead, the oscillations might be general signs of neuronal activation that co-occur when neurons are excited and information is being processed, but have little or no specific informational function. For example, gamma rhythms in cortical populations are reflections of excitatory and inhibitory dynamics of pyramidal and basket cells that appear when cortical pyramidal cells are maximally driven, but there appears to be little or no information conveyed in specific oscillatory frequencies. In some cases stimulus detection thresholds are lower when stimulus presentations are timed to coincide with recovery phases of oscillations, but this may simply reflect the larger numbers of neurons available and ready to respond at those moments. Here oscillations play a somewhat more tangential, facilitating role vis-à-vis neural coding and information processing.

Perhaps the field can learn from its history. In the past an intriguing "alpha scanning" mechanism was proposed as a substrate for computing form invariants (McCulloch, 1951), but this hypothesis was severely undermined by the relative ease that alpha rhythms can be disrupted at will without major perceptual or cognitive consequences. Today critical experiments likewise need to determine whether phase-resets or abolition of oscillations using appropriately timed stimuli, such as clicks, flashes, shocks, or pharmaceutical interventions can significantly disrupt functions. Experiments along these lines could clarify what dependencies exist between neural information processing mechanisms and the stimulus-driven, stimulus-triggered, and intrinsic oscillatory neurodynamics of neuronal excitation, inhibition, and recovery.

13.3.5.3 Time-Delay Neural Networks, Synfire Chains, and Timing Nets

Thus far, both traditional connectionist networks and synchronized, oscillatory, and/or temporally gated connectionist network assume channel coding of specific stimulus attributes. In the early auditory system, however, many stimulus distinctions appear to be conveyed by means of temporal codes.

Time-delay neural networks can be used to interconvert time and place (channel) patterns. In essence, any fixed spatiotemporal spike volley pattern can be recognized and produced by implementing appropriate offsetting time delays within and/or between neural elements. Classical time-delay networks used systematic sets of synaptic and axonal transmission delays embedded in arrays of coincidence detectors to convert temporal patterns to activations of specific channels. These include temporal correlation models for binaural localization (Jeffress, 1948), periodicity pitch (Licklider, 1959), and binaural auditory scene analysis (Cherry, 1961).

Modulation-tuned elements can be also used to convert time to place, and periodotopic maps consisting of such elements have been found in the auditory pathway (Schreiner & Langner, 1988). These maps form modulation spectrum representations of periodicities below 50 Hz that can usefully subserve recognition of consonantal speech distinctions and rhythmic patterns. Although neural modulation spectra have been proposed as substrates for periodicity pitch, modulation-based representations for pitch break down when confronted with concurrent harmonic tones (e.g., two musical notes a third apart).

Synfire chains (Abeles, 2003) and polychronous networks (Izhikevich, 2006) are time-delay networks in which spatiotemporal channel activation sequences are propagated. These are distinct from both connectionist and time-delay networks in that both channel and timing are equally important. Information is encoded in the spatiotemporal trajectory of spikes through the system. Because each trajectory depends on specific interneural delays and synaptic weightings, it is unclear how stimulus invariances and equivalences might be realized this way. However, one of the major potential advantages of synfire and polychronous networks is their ability to multiplex signals. In these networks a given neuron can participate in multiple synfire chains and polychronous patterns, and this mutual transparency of signals drastically simplifies the neurocomputational problems of representing multiple attributes and objects.

Timing nets are a third general type of neural network that are distinct from both connectionist networks and time-delay networks (Cariani, 2001a, 2004). Whereas connectionist networks operate entirely on channel activation patterns, and time-delay networks convert temporal patterns into channel activations, timing nets operate entirely in the time domain. Timing nets are similar to time-delay neural networks in that they consist of arrays of coincidence detectors interconnected by means of time delays and synaptic weights. Whereas both types of networks have temporally coded inputs, the outputs of timing nets are also temporally coded rather than by channel.

Simple timing nets have been proposed for analysis of periodicity and spectrum and for grouping and separation of auditory objects. Feedforward timing nets act as temporal pattern sieves to extract common spike patterns among their inputs, even if these patterns are interleaved with other patterns. Such operations elegantly extract common periodicities and low-frequency spectra from two signals, for example, recognizing the same vowel spoken by two speakers with different voice pitches (different fundamental frequencies [F0s], same spectra) or different vowels spoken by the same speaker (different spectra, same fundamental frequencies). Such networks can also be used to separate out and recognize embedded and interleaved temporal patterns of spikes, an important property for multiplexing of multiple temporal pattern signals and for complex, multidimensional temporal representations. Timing nets illustrate how processing of information might be achieved through mass statistics of spike correlations rather than through highly specific connectivities.

Recurrent timing nets consist of delay loops and coincidence elements that carry circulating temporal patterns associated with a stimulus (Cariani, 2001a, 2004; see also the recurrent neural loop model of Thatcher & John, 1977). The nets in effect multiply a signal by its delayed version to build up and separate multiple repeating temporal patterns that are embedded in the signal. The auditory system readily separates multiple musical notes whose fundamental frequencies (F0s) are separated by more than 10% (e.g., nonadjacent notes on the piano). Such note combinations have embedded within their waveforms two different patterns that have different repetition times (fundamental periods). The time-domain filtering operations carried out by the delay loops act roughly like comb filters to produce two sets of signals that resemble the individual vowel waveforms. In neural terms, they separate the two vowels on the basis of invariant temporal patterns of spikes rather than by segregating and binding subsets of activated periodicity or spectral feature channels. In doing so, they provide an example of how auditory object formation based on harmonic, periodic structure could occur at very early stages of auditory processing, before any explicit frequency and periodicity analysis takes place. On larger timescales, such networks can build up and separate repeating, complex rhythmic patterns as well (Cariani, 2002).

Feedforward and recurrent timing nets were developed with temporal coding of pitch and auditory scene analysis in mind. Because they operate on temporal patterns of spikes that are not evident at the level of auditory cortex, neural timing net mechanisms for periodicity pitch analysis and F0-based sound separation would likely need to be located at earlier stages of auditory processing, possibly dynamically

facilitated by descending projections to thalamus and midbrain (see discussion of reverse-hierarchy theory in Section 13.3.6.2). Because coarser temporal patterns of spikes associated with onsets and offsets of auditory events are present in cortical stations, recurrent timing mechanisms could exist at those levels to carry out coarser temporal pattern comparisons whose violations produce mismatch negativities.

13.3.6 Functional Roles of the Auditory Cortex

In considering the functional role of the human auditory cortex vis-a-vis the rest of the brain, it is useful first to summarize some general principles that govern brain organization and function.

13.3.6.1 General Principles of Brain Organization and Function

In cybernetic terms, brains can be seen as adaptive, goal-directed percept-action systems. Sensory systems gather information about the surrounding world (sensory functions). Cognitive representations and operations evaluate incoming sensory inputs and prospective actions in the context of previously acquired knowledge. Motor systems carry out actions on the world (motor functions). Coordinative linkages, from simple reflex arcs to much more complex circuits, link percepts and cognitive representations to actions. Motivational goal systems steer perception and action toward satisfaction of immediate needs, while anticipatory and deliberative systems analyze the deeper ramifications of sensed situations and plan prospective actions (executive functions) that satisfy longer range goals. Evaluative reward systems judge the effectiveness of sensorimotor linkages vis-à-vis goals and adaptively modify neural subsystems to favor behaviors that fulfill drive goals to avoid those that are detrimental to survival. Affective and interoceptive systems provide a running estimate of the state of the organism that influence choice of behavioral alternatives (e.g., fight/flight). Mnemonic systems retain associations between sensory information, internal deliberations, sensorimotor sequences, and rewards for later use by steering mechanisms that take into account anticipated consequences of action alternatives (rewards and punishments).

These different functionalities are subserved by different subcortical and cortical neuronal populations (Mesulam, 2000). Cerebral cortical regions are involved in sensory, motor, coordinative sequencing, anticipatory, and executive functions. The cerebellum involves real-time motor adjustments and control of sensory surfaces. Hypothalamus and amygdala are involved with fixed drives and affect-based modulation of behavior. Dopaminergic predictive reward circuits reconfigure the system to incorporate new goals. Basal ganglia structures steer attention and switch action modes to address current, salient goals, providing linkages between limbic-generated goal states and cortical sensorimotor processing.

Some basic principles exist for cortical organization. Within general neuroanatomical plans that are specified through genetic guidance of developmental processes, most large-scale patterns of cortical functional connectivity can be understood

through the interaction of correlated external inputs, internal reward signals, existing interneural connectivities, and the action of activity-dependent biophysical mechanisms that alter them.

The first maxim is "cortex is cortex," meaning that different cortical regions have roughly the same cell types and general organization, albeit with varying relative cell densities and connectivities among and within the cortical layers. A second is that the "stimulus organizes the tissue" such that the dominant inputs to a given region alter the fine structure and function of the tissue according to the correlational structure of its inputs and outputs vis-à-vis effective action. The functional organization of unimodal cortex is largely determined by the afferent inputs and ultimately by the organization of sensory and motor surfaces. Thus, auditory cortex has several fields that are coarsely cochleotopically organized, in parallel with retinopic organization in visual cortex, and somatotopic organization in somatosensory cortex.

A third organizing principle is that there is an ongoing competition for cortical territory that is mediated by the strength of both incoming information and internal evaluative reward signals. The strongest, most internally rewarded inputs come to dominate the responses of a given region over time. When normal sensory inputs to a patch of cortex are silenced, other weaker inputs are strengthened (by sprouting and synaptic proliferation, stabilization, and strengthening). Provided that they play a useful functional role such that they are internally rewarded, such weak inputs can then come to dominate responses.

A fourth rule-of-thumb is that connectivities between neural populations are almost invariably reciprocal, such that recurrent loops are norm rather than exception. "Everything is connected" by such recurrent loops, that is, there are multisynaptic pathways that provide reciprocal connections between any two neurons in the system. Because "neurons that fire together wire together" even arbitrary long-range reciprocal connections can be made and stabilized. Lastly, lateral interconnections are mostly local and short range. These connectivity patterns lead to cortical convergence zones that handle confluences of different types of sensory information (Damasio & Damasio, 1994), provided that the different types of information correlate in a functionally meaningful way (internally rewarded). Cortical regions that operate on similar kinds of information and/or perform similar tasks therefore tend to be clustered together spatially. Much of the large scale functional topography of cortical regions may ensue from these basic principles (e.g., dorsal paths for localization leading to body and extrapersonal space maps in the parietal lobe, ventral paths for object recognition leading to regions in the temporal lobe, hemispheric colocalizations of related, time-critical functions).

13.3.6.2 Conceptions of Auditory Cortical Function

The auditory cortex receives incoming sensory information from the ears via ascending afferent auditory pathways, and controls the information it receives through descending, efferent pathways that modulate neural activity at every level of processing (Fig. 13.5; Winer, 1992; Clarke and Morosan, Chapter 2). The auditory

cortex has reciprocal connections with other cortical regions involved with object recognition and classification (temporal lobe), analysis and production of sensorimotor sequences (premotor frontal regions), expectancy and decision making (frontal regions), body space (parietal regions), as well as with uni- and multimodal cortical regions associated with other sensory systems.

By virtue of its connections to the auditory pathway and to other functionally related cortical and subcortical (limbic, basal ganglia) areas, auditory cortex is strategically situated to coordinate processing of auditory information for a number of organism-level purposes. These purposes include monitoring changes in the environment (alerting functions), separating sound objects and streams (perceptual organization), detecting and discriminating relevant sounds (discriminatory functions), recognizing familiar sounds (classificatory and mnemonic functions), locating relevant sound sources (orienting functions), decoding speech communication signals (phonetic, syllabic, and word classification and sequence analysis functions), providing feedback for sound production processes, and self-regulation of internal state (e.g., use of music to regulate mood, affect, pleasure, arousal).

Currently two broad conceptions exist concerning the role of auditory cortex vis-à-vis lower stations (perspectives often heavily shaped by whether one has investigated the system at subcortical or cortical levels). The first conceives of auditory cortex as the culmination of the auditory pathway, the stage at which all incoming auditory information is organized and analyzed. Here auditory cortex is the nexus for fine-grained representations of sound that are used for auditory functions. In this sequential-hierarchical feedforward view, it has been assumed that "higher level functions" such as recognition of phonetic tokens and the organization of the auditory scene take place at the cortical level after a basic frequency and spatial hearing analysis has been first carried out by lower stations.

A second, emerging perspective conceives of auditory cortex as a control system. In vision this has been termed "the reverse hierarchy theory" (Ahissar & Hochstein, 2004). The main purpose of such a control system is not as a repository of fine-grained representations. Rather, it is to organize information processing in "lower" circuits at thalamic, midbrain, brain stem, and even perhaps cochlear levels by means of descending connections that can release inhibitory controls. This disinhibitory control may be similar in function to the double-inhibitory mechanism by which in basal ganglia activity release inhibition to bias activity patterns in cortical motor areas toward particular actions (movement initiation, switching) and to bias sensory areas to facilitate particular signals (attention). The system in effect chooses its own inputs contingent on its immediate interests.

The representations needed for such a control system do not necessarily need to be as precise as perceptual acuities if the cortex can access fine-grained temporal information at lower stations when needed. When presented with a task requiring attention and fine discrimination, the cortex could potentially pose the question to lower levels by setting up (by disinhibition) dynamic neural linkages that facilitate and hold the informational distinctions that are needed. This theory has the merits that it is consistent with the massive descending pathways that are present in both the auditory and visual systems, and it also provides some explanation as to how

fine-grained temporally coded information might be used by central stations in the auditory system, yet not be present in precise and overt form. It is consistent with relatively recent evidence that cortical activity may modulate lower level processing even as far down as the brain stem, on both short- and long-term timescales (Tzounopoulos et al., 2004; Lee et al., 2009).

13.4 Fundamental Issues and Open Problems

13.4.1 Identifying Neural Codes and Representations at the Cortical Level

Perhaps the most fundamental open problem at the cortical level is to identify the specific neural codes that subserve different perceptual and cognitive representations, such as pitch, timbre, location, and loudness (Phillips et al., 1994; Brugge et al., 1996; Furukawa & Middlebrooks, 2004; Bendor & Wang, 2005; Bizley & Walker, 2010; Hall and Barker, Chapter 7). Cortical representations related to pitch and rhythmic pattern are most important for music (Zatorre and Zarate, Chapter 10), whereas those related to timbral, phonetic distinctions are most important for speech communication (Huetz et al., 2011; Giraud and Poeppel, Chapter 9). The nature of cortical codes places strong constraints on neural mechanisms for higher-level informational integration, in the specific processes that form auditory objects and streams (Shamma & Micheyl, 2010; Griffiths, Micheyl, and Overath, Chapter 8), in the integration of auditory representations with those of other senses (van Wassenhove and Schroeder, Chapter 11), and in the utilization of auditory information for action (Hickok and Saberi, Chapter 12). This section lists and briefly describes some of the most important unresolved issues concerning the nature of auditory codes and representations that apply generally to all of the aforementioned problem domains of basic auditory constituents, music, speech, auditory scene analysis, multimodal representations, and sensorimotor integration.

13.4.1.1 Rate, Channel, and Time Codes

Because of their prominent and abundantly documented tonotopic organization, the peripheral and central auditory systems have often been conceived as an ensemble of labeled-line frequency channels, such that profiles of average firing rates across tonotopic axes provide a central, general-purpose representation of the stimulus power spectrum. Similarly, cortical units whose average firing rates covary with many other acoustic parameters, such as periodicity, intensity, duration, amplitude and frequency modulation, bandwidth, harmonicity, and location have been found. This leads to the hypothesis that representation of the auditory scene at the cortical level is simply a matter of analyzing average firing rate profiles among a relatively

small number of neural subpopulations that encode feature maps. Such coding schemes work best with elements that have stable receptive fields, with sensitivity to only one or two acoustic parameters. Complicating this picture, however, is the problem of disentangling the multiple parameters that can influence any given neuron's firing rate, especially if multiple auditory objects are simultaneously present.

A strong case can be made that the central representations for both periodicity pitch and spectral determinants of timbre are ultimately based on population-wide interspike interval statistics at early stages of auditory processing (Palmer, 1992; Cariani & Delgutte, 1996; Cariani, 1999; Ando & Cariani, 2009). Although individual neurons and neuronal ensembles in lightly and unanesthetized auditory cortex can phase lock up to stimulus periodicities of several hundred Hz (Fishman et al., 2000; Wallace et al., 2002), most cortical neurons do not go above 30–40 Hz (Miller et al., 2001). Thus, the direct, iconic temporal-pattern codes for pitch and timbre that predominate in the auditory periphery and brain stem appear to be largely absent at the cortical level, necessitating some form of coding transformation (Wang, 2007). The most specific neural correlates of pitch found to date in auditory cortex instead involve specialized subpopulations of neurons whose firing rates are tuned to particular periodicities (Bendor & Wang, 2005). Questions of how peripheral timing patterns might be transformed in the central auditory system to give rise to such cortical pitch detectors are still unresolved.

However, other types of temporal codes that are based on the relative latencies of spikes rather than stimulus-driven temporal patterns are possible at the cortical level. Neurons in A1 appear to encode stimulus onset timing very precisely in their response latencies (Heil, 1997; Phillips et al., 2002). Representations can be based on latency differences across units (i.e., latency-place) codes, or dynamic latency-coding schemes (Heil, 1997). For example, the loudness of an abrupt, short duration tone can be encoded by the temporal dispersion of first-spike responses over a population. Multiplexed sparse distributed temporal codes (Abeles et al., 1993; Villa, 2000; Panzeri et al., 2009) in which periodicity-related spikes are interspersed with those encoding other kinds of perceptual information (timbre, spatial attributes) may exist in auditory cortex (Chase & Young, 2006) in some covert form that is difficult to recognize. Some evidence exists for precise temporal sequences of spikes that are related to perceptual functions (Villa, 2000). Because the latency of these sequences can vary from trial to trial, they may be smeared in poststimulus time histograms.

13.4.1.2 Sparse-Efficient versus Abundant-Redundant Codes

With the advent of information theory and its application to neuroscience and psychology, the degree of redundancy of neural responses at various levels of processing within sensory system has become a key issue in the analysis of neural representations. Horace Barlow proposed that neural representations of stimuli become less and less redundant at each successive processing stage within sensory systems (Barlow, 1961). In this context, the question of whether representations at

the level of auditory cortex are in some sense less redundant than the representations at lower levels of the auditory system has been raised (Chechik et al., 2006).

Another important question related to coding redundancy concerns the "sparseness" of neural representations. One way to characterize sparseness involves counting how many neurons in a population are active during the presentation of a stimulus, and how many are quiescent (Hromadka et al., 2008, Bizely et al., 2010). If only a relatively small number of neurons are active (e.g., <10%), the neural representation of the considered stimulus is said to be sparse. Another approach to sparseness involves counting how many spikes each neuron produces in response to a stimulus. In theory, sparse representations are desirable because they are more energetically efficient. The downside, of course, is reduced resilience to individual-component failure, or malfunction.

If sound representations in auditory cortex are efficient, and sparse, one may wonder why there should be so many more neurons at the cortical level, compared to lower stations in the auditory system. One possible answer to this question is that auditory cortex has many other functions besides the efficient representation of sound. In particular, it may have to perform complicated computations on multiple auditory representations that in turn need to be registered and coordinated with information provided by other sensory modalities (DeWeese et al. 2005, van Wassenhove & Schroeder, Chapter 11). Several recent studies have identified neurons in auditory cortex whose responses are modulated by nonauditory influences (Bizley & King, 2008; Kayser et al., 2008; Panzeri et al., 2009).

While questions of how and to what extent the redundancy of neural representations vary as one ascends the auditory pathway, perhaps the more fundamental question is why this should be so in the first place. From a functional point of view, lower redundancy makes for more efficient coding in an information-theoretic sense. On the other hand, in the face of abundant sources of both internal and external noise, redundancy also plays a critically important role in enhancing reliability. Therefore, one would expect a well designed neural information–processing system to achieve a judicious balance between efficiency and redundancy.

It is possible that Barlow's coding hypothesis is not testable given our current level of understanding of neural coding at the cortical level. A pervasive problem with optimality arguments in biology is that one does not know a priori for what specific functions the system has been optimized, and what constraints (structural, developmental, evolutionary) have shaped it. Optimality analysis will rest on much firmer ground once the basic operating principles of the system (codes, computations, functions) are better understood and various design trade-offs can be more realistically assessed.

13.4.1.3 Coding of Features versus Objects

Neurons in primary and secondary auditory cortex have been found to respond in a selective manner to various sound "features," such frequency sweeps (Tian & Rauschecker, 1994), bandwidths (Rauschecker & Tian, 1994), or temporal and/or

spectral modulation rates (Kowalski et al., 1996). However, many of these features are already extracted and represented in some way in lower stages of the auditory system. Thus, even though some important differences have been identified between cortical and lower-level responses (e.g., in the broadness of tuning, the nonmonotonicity of rate-level functions), it is tempting to think that there must be more to auditory cortex function than just the extraction and representation of disjoint features. This leads to the notion that auditory cortex may be a place where representations of various sound features are conjoined in a meaningful way to form representations of auditory objects (Nelken et al., 2003). Empirical evidence for the representation of auditory objects, or streams, and not just features at the level of auditory cortex, however, still remains very limited. One line of evidence comes from the results of several single-unit electroencephalography (EEG), magnetoencephalography (MEG), and functional magnetic resonance imaging (fMRI) studies, which concur to indicate that neural responses in primary and/or secondary auditory cortex reflect auditory streams (Shamma & Micheyl, 2010; Shamma et al., 2011; Griffiths, Micheyl, and Overath, Chapter 8). Another line of evidence that neural responses in auditory cortex reflect not just physical stimulus properties, but also the perceptual organization of these features into objects, comes from EEG studies that have identified a wave (the "object-related negativity"), which appears to depend specifically on whether listeners hear out a mistuned component in an otherwise harmonic complex as a separate object (Alain and Winkler, Chapter 4). Although these findings provide important hints that auditory cortex does indeed represent auditory objects, additional research is needed to clarify the neural mechanisms whereby representations of different sound features are combined to form representations of auditory objects at the level of auditory cortex.

13.4.2 The Hyperacuity Problem

For many perceptual discriminations, the most highly tuned receptive fields of neural elements are typically much coarser—by one to two orders of magnitude than the finest distinctions that can be made by the organism as a whole. The problem of accounting for this apparent discrepancy, which exists in nearly every sensory modality, is known as the hyperacuity problem (Rieke et al., 1997). A striking example of hyperacuity problem in the auditory modality relates to the relationship between neural frequency selectivity and behavioral frequency discrimination. Just-noticeable differences (JNDs) in the frequency of moderate-level pure tones below 2 kHz can be as small as 0.1–0.2% (Moore, 1973). At 1 kHz, this corresponds to a frequency difference of about 1 Hz. In contrast, at moderate sound levels the firing-rate response bandwidths of auditory neurons at all stations in the pathway are typically on the order of large fractions of an octave (Evans et al., 1992). Although there have been recent reports of "ultrafine" frequency tuning at the single-unit level in human auditory cortex (Bitterman et al., 2008), the frequency JNDs that were estimated based on such tuning (around 3%) are still an order of magnitude larger than

the smallest JNDs that can be achieved by human listeners. The usual "solution" to this discrepancy assumes that the latter do not rely on rate-place representations, but rather on temporal information, that is, phase locking. However, because phase-locking decreases sharply at successively higher auditory stations (Cariani, 1999), this type of explanation is unlikely to apply at the level of auditory cortex. Thus, either behavioral frequency discrimination is determined below cortex, or sufficiently precise neural representations of pure-tone frequency must exist at the level of auditory cortex that can account for the exquisitely small JNDs that are observed in humans and other animals.

13.4.3 The Invariance Problem

Typically, many auditory attributes can be highly invariant with respect to changes in sound parameters. A prime example is perceptual invariance of low-frequency sounds with respect to stimulus intensity. Although the loudness of sounds invariably increases monotonically as a function of sound pressure level, for low-frequency sounds the same sound presented at different levels is recognizably similar in pitch, timbre, duration, and location. For high-frequency tones, however, pitch and timbre are much more labile.

These perceptual invariances are obtained despite profound changes in both absolute and relative neural firing rates at all levels of auditory processing. Cortical sound-responsive neurons with nonmonotonic rate-level functions are quite common, which greatly complicates population-based explanations of level-invariant percepts and equivalence classes (Tramo et al., 2005). It is one of the main reasons that coherent rate-based tonotopic spatial organization is only seen only at low sound pressure levels near neural response thresholds and breaks down at higher levels (Phillips et al., 1994). Ironically, level-invariant, rate-based frequency tunings have been observed in marmoset cortex for high-frequency pure tones (Sadagopan & Wang, 2008), the very stimuli for which human pitch percepts are the least invariant with respect to level.

A second example of invariance is the relative stability of pitch, timbre, loudness, and location with respect to sound duration. This stability generally holds for durations longer than 50–100 ms. For shorter time periods, pitch strength, timbre, and loudness can change dramatically with duration. A third major invariance is the relative stability of pitch, timbre, loudness, and duration with respect to sound-source location relative to the listener.

Related to perceptual invariances are perceptual equivalence classes. Sounds consisting of low-frequency, resolved harmonics that have different phase spectra (and consequently waveform envelopes) nevertheless are indistinguishable. Harmonic, low-frequency sounds having the same fundamental frequency almost invariably produce the same low pitch at the fundamental, despite profound differences in spectral content. Pitch equivalence classes are especially important in music, where various instruments with differing spectral and dynamic characteristics play

the same notes that evoke the same pitches. This pitch equivalence is what permits different types of instruments to readily serve as tuning references for each other. Octave equivalences produce pitch chromas that form the foundation for tonal pitch classes in music theory. That these same broad pitch equivalence classes extend to fundamental frequencies well beyond the range of human voices, and that they are shared by a phyletically broad range of animal listeners strongly suggests that they are integral products of basic auditory mechanisms for analysis and separation of sounds rather than the products of ontogenetic associative learning or recent evolution. At the level of the auditory nerve, pitch and octave equivalence falls out of common features in all-order interspike interval codes (Cariani & Delgutte, 1996; Cariani, 1998, 2002), whereas at the cortical level pitch equivalence may be manifested by the responses of periodicity-tuned neurons (Bendor & Wang, 2005).

13.4.4 The Transformation Problem

In vision, within limits, shapes remain perceptually invariant, such that they can be recognized when translated, rotated, and magnified with respect to retinal coordinates. This was known to the Gestaltists as form invariance under transformation. Despite the large changes that occur in retinotopic patterns of activity, the representations of these shapes nevertheless retain essential, relational aspects that are used to judge similarity and to support recognition. In audition and the temporal sense, three analogous invariances exist for pitch relations, timbral relations, and temporal event relations. These are, respectively, transpositional invariance for melodies and chords, timbral invariance of for vowels spoken by different speakers, and tempo invariance for rhythmic patterns.

Melodies are temporal sequences of pitched-events. Transpositional invariance is illustrated by the common observation that musical melodies can be identified even after they are transposed into a different key or register (frequency range). Transpositional invariance involves the ability to recognize a melody on the basis of relative pitch relations, irrespective of the absolute fundamental frequency of the beginning note. The operation of transposition multiplies all frequencies by a constant factor, thereby retaining the same frequency ratios and proportionalities. Recognition of transposed melodies is highly reliable if the melody is familiar and/or harmonically well structured (i.e., "tonal"), and transposed notes all bear the same frequency ratios (i.e., in musical terms, if musical intervals are preserved), but is much weaker and conditional if only pitch contours (patterns of up–down transitions of successive pitches) are retained (Handel, 1989; McDermott & Oxenham, 2008).

Chords can also be transposed. Chords are multiple notes played together. The type of a chord (e.g., major vs. minor vs. diminished or augmented) is determined by the musical intervals (frequency ratios) between its constituent notes. With a little exposure, human listeners can distinguish different types of consonant and dissonant chords irrespective of the absolute note frequencies that constitute them. The existence region for transpositional invariance of melodies and chords

parallels that for musical tonality. Transpositional invariance, being based on musical intervals, appears to be associated with periodicity pitch, and may therefore ultimately depend on properties of temporal, interspike interval codes for periodicity pitch in early auditory processing.

Timbral invariance involves ability to recognize common timbral qualities despite changes in absolute acoustical parameters. Perception of phonetic distinctions in speech is relatively invariant with respect to the considerable acoustical variations that are produced by different speakers with different vocal tract sizes. In early studies of vowels, phoneticists found that male and female productions of the same, perceptually equivalent vowels have different absolute formant frequencies, but relatively more similar formant ratios. Interestingly, sensitivity to formant ratios has recently been observed in MEG responses to synthetic vowels in auditory cortex (Monahan & Idsardi, 2010). Vowel normalization is an operation that produces a more invariant representation by taking into account formant ratios (F2/F1, F3/F2, F3/F1) and/or formant-voice pitch ratios (F1/F0, F2/F0, F3/F0). In the auditory nerve, the most intense harmonic in each formant region dominates the interspike intervals that are produced ("synchrony capture"), such that the temporal representation of vowels resembles that produced by a small number of harmonically related pure tones (Delgutte & Kiang, 1984). The formant frequency ratios that may determine the different timbral categories of vowels are thus not unlike the tonal frequency ratios that constitute different musical intervals and chords (see also the timbral intervals discussed in McAdams & Giordano, 2009). Thus, similar kinds of mechanisms conceivably subserve the transpositional invariances of musical intervals, chords, melodies, and even vowel timbres.

Tempo invariance involves the ability to recognize a rhythmic or melodic pattern when played at different speeds. As long as the time intervals between notes are neither too short nor too long (roughly, $0.1\ s < I < 2\ s$), the temporal pattern invariance holds as long as the time intervals are all changed proportionately.

Invariance under transformation is a fundamental unsolved problem for computational neuroscience (von der Malsberg, 1994; Wiskott, 2006). In the late 1940s, Pitts and McCulloch proposed neural networks to carry out both visual (translation, magnification) and auditory (melodic transposition) transformations (Pitts & McCulloch, 1947; McCulloch, 1951). Their representational model used diagonally crossing sets of projections on logarithmic retinotopic and cochleotopic cortical place maps to implement "shifter" circuits that would recognize angle and frequency ratios. However, if the underlying neural representations instead involve temporal patterns of spikes, then time-warping of these patterns, that is, stretching or compressing time intervals by a constant factor, can provide a general solution to the three auditory invariances (Boomsliter & Creel, 1962). The different temporal regimes associated with the three transformations would likely require processing at different levels. Fine-grained temporal information needed for recognizing harmonic ratios for recognitions of musical intervals and vowels is ubiquitous in early auditory stations, whereas coarse-grained temporal information for recognizing rhythmic patterns of events also exists over large portions of cerebral cortex (Thatcher & John, 1977).

13.4.5 Temporal Integration and Auditory Memory Mechanisms

Processing of sounds and sound sequences occurs over different time regimens that span echoic memory integration windows for pitch and timbre, and loudness summation, intermediate duration windows for melodic and rhythmic pattern integration, and still longer temporal windows for large-scale recurring patterns (Snyder, 2000, 2009; Trainor & Zatorre, 2009). When performing sequential matching tasks, human listeners can easily hold precise memories of pitch, timbre, loudness, location, and other auditory qualities for several seconds provided that subsequent distractions do not intervene (Demany & Semal, 2007). In musical contexts, tonal and rhythmic expectations can persist over even longer durations (Patel, 2008). To appreciate the complex interplay of multiple memory processes, one has only to think of an extended piece of tonal symphonic music, with its many excursions to and from tonal centers, metrical frames, and melodic motifs (Bigand, 1993).

The nature and locations of the various memory traces remain to be identified (Fritz et al., 2005), and their workings likely depend on the nature of the neural codes that are involved. For example, rate-place codes might entail persistently active subsets of neurons that encode particular features, whereas temporal codes might utilize reverberatory circuits that maintain temporal patterns of activity over time. Adaptation of neural responses over different timescales (ranging from milliseconds to several tens of minutes) likely plays an important role in the representation of temporal sound sequences in auditory cortex (Ulanovsky et al., 2004), and may potentially explain many aspects of music and speech perception.

13.4.6 Neural Requisites for Conscious Auditory Awareness

A great deal of progress has been made in the scientific study of the neural basis of consciousness over the last decade. The best current theories of the neural requisites of awareness involve the necessity of recurrent activation patterns for a given stimulus to become supraliminal (Lamme, 2006). Currently there is debate about whether recurrent corticocortical or thalamocortical activation of modality-specific pathways are sufficient (albeit without the ability for overt report), or whether recurrent activation patterns need also to include frontal and/or parietal regions as well. Recurrent activation of frontal regions results in systemic recurrence for support of global workspaces, while parietal activation of body/self maps may be essential for "ownership" of percepts (Pollen, 2008) or for providing a requisite level of attentional gain through associated basal ganglia circuits. Recurrent activation may facilitate attainment of a threshold degree of informational complexity (Tononi & Koch, 2008) or it may support dynamic regeneration of neuronal signals necessary for supporting sustained, stable systemic informational states in the first place (Cariani, 2000).

The vast majority of neurophysiological and psychophysical studies have involved visual experience, but any truly general theory of the neuronal basis for awareness

needs to apply to other kinds of sensory experience as well. This makes the auditory system an ideal testing ground for theories developed using examples from vision. In the last decade striking auditory analogues to visual neglect syndromes and blindsight have been reported (Garde & Cowey, 2000, Clarke & Thiran, 2004). As in vision, it appears that body space representations in the parietal lobe must be engaged for auditory percepts to enter awareness, and also that the presence of auditory stimuli can be detected in the absence of direct experience of their qualities.

Many general and specific hypotheses concerning consciousness await investigation by auditory scientists. Is recurrent activation of frontal supramodal regions either essential or sufficient for auditory experience? Does conscious auditory awareness of an external sound event require completion of frontal–temporal feedback loop? Practically, to understand central tinnitus, one wants to identify the requisites for an endogenously generated neural pattern of activity to become part of conscious awareness. Beyond restoring auditory discriminatory capacities, it is also desirable to restore the subjective, felt texture of hearing in those who have lost or never had it, for example, the restoration of the experienced sound qualities of speech and music in cochlear implant users. Here a neurophenomenology that surveys the gamut of auditory experiences and identifies their neural correlates is a prerequisite. Whether in pursuit of restorative therapies or basic knowledge, auditory neuroscience will eventually develop such a neurophenomenological theory that will finally bridge the divide between our brains and our auditory experiences to provide useful and meaningful answers to fundamental questions of what and how we hear.

13.5 Summary

Although biological brains are impressively powerful informational engines, they are neither omnipotent nor infinitely complex—and there is no reason to believe that they cannot be understood by human minds properly equipped with the right conceptual tools. If the information functions of auditory cortex are to be understood, neurocomputational theories and neurophysiological experiments need to pay close attention to and strive to explain the large-scale structure of auditory perception and cognition. Because not all aspects of cortical structure and neural activity necessarily play critical roles in its informational functions, it is therefore essential that the cortical neural codes that do play such roles be identified as early as possible. As with the elucidation of the genetic code half a century ago, once the signals of the system are identified, understanding of the rest of the functional framework should quickly follow.

Acknowledgments This work was supported by research grants from the Air Force Office for Sponsored Research (FA9550-09-1-0119 to P. C.) and the National Institutes for Health (R01 DC05216 and R01 07657 to C. M.).

References

Abeles, M. (2003). Synfire chains. In M. A. Arbi (Ed.), *The handbook of brain theory and neural networks*, 2nd ed. (pp. 1143–1146). Cambridge, MA: MIT Press.

Abeles, M., Bergman, H., Margalit, E., & Vaadia, E. (1993). Spatiotemporal firing patterns in the frontal cortex of behaving monkeys. *Journal of Neurophysiology*, 70, 1629–1638.

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Science*s, 8, 457–464.

Ando, Y., & Cariani, P. (2009). *Auditory and visual sensations*. New York: Springer.

Barlow, H. B. (1961). Possible principles underlying the transformation of sensory messages. In W. A. Rosenbluth (Ed.), *Sensory communication* (pp. 217–234). Cambridge, MA: MIT Press.

von Bekesy, G. (1967). *Sensory inhibition*. Princeton, NJ: Princeton University Press.

Bendor, D., & Wang, X. (2005). The neuronal representation of pitch in primate auditory cortex. *Nature*, 436, 1161–1165.

Bigand, E. (1993). Contributions of music to research on human auditory cognition. In S. McAdams & E. Bigand (Eds.), *Thinking in sound: The cognitive psychology of human audition* (pp. 231–277). Oxford: Oxford University Press.

Bitterman, Y., Mukamel, R., Malach, R., Fried, I., & Nelken, I. (2008). Ultra-fine frequency tuning revealed in single neurons of human auditory cortex. *Nature*, 451, 197–201.

Bizley, J. K., & King, A. J. (2008). Visual-auditory spatial processing in auditory cortical neurons. *Brain Research*, 1242, 24–36.

Bizley, J. K., & Walker, K. M. M. (2010). Sensitivity and selectivity of neurons in the auditory cortex to the pitch, timbre, and location of sounds. *The Neuroscientist*, 16, 453–469.

Bizley, J. K., Walker K. M., King, A. J., & Schnupp, J. W. (2010). Neural ensemble codes for stimulus periodicity in auditory cortex. *Journal of Neuroscience*, 30(14), 5078–5091.

Boomsliter, P., & Creel, W. (1962). The long pattern hypothesis in harmony and hearing. *Journal of Music Theory*, 5, 2–31.

Boring, E. G. (1942). *Sensation and perception in the history of experimental psychology*. New York: Appleton-Century-Crofts.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.

Brugge, J. F., Reale, R. A., & Hind, J. E. (1996). The structure of spatial receptive fields of neurons in primary auditory cortex of the cat. *Journal of Neuroscience*, 16, 4420–4437.

Bullock, T. H. (1992). Introduction to induced rhythms: A widespread heterogeneous class of oscillations. In E. Basar & T. H. Bullock (Eds.), *Induced rhythms in the brain* (pp. 1–26). Boston: Birkhauser.

Buzsáki, G. (2006). *Rhythms of the brain*. New York: Oxford University Press.

Cariani, P. (1995). As if time really mattered: Temporal strategies for neural coding of sensory information. *Communication and Cognition–Artificial Intelligence (CC-AI)*, 12, 161–229. Reprinted in K. Pribram (Ed.), *Origins: Brain and self-organization* (pp. 208–252). Hillsdale, NJ: Lawrence Erlbaum.

Cariani, P. (1999). Temporal coding of periodicity pitch in the auditory system: An overview. *Neural Plasticity*, 6, 147–172.

Cariani, P. (2000). Regenerative process in life and mind. In J. L. R. Chandler & G. Van de Vijver (Eds.), *Closure: Emergent organizations and their dynamics*. *Annals of the New York Academy of Sciences*, 901, 26–34.

Cariani, P. (2001a). Neural timing nets. *Neural Networks*, 14, 737–753.

Cariani, P. (2001b). Temporal coding of sensory information in the brain. *Acoustic Science & Technology*, 22, 77–84.

Cariani, P. (2002). Temporal codes, timing nets, and music perception. *Journal of New Music Research*, 30, 107–136.

Cariani, P. (2004). Temporal codes and computations for sensory representation and scene analysis. *IEEE Transactions on Neural Networks*, *Special Issue on Temporal Coding for Neural Information Processing*, 15, 1100–1111.

Cariani, P., & Delgutte, B. (1996). Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. II. Pitch shift, pitch ambiguity, phase-invariance, pitch circularity, and the dominance region for pitch. *Journal of Neurophysiology*, 76, 1698–1734.

Carr, C. E. (1993). Processing of temporal information in the brain. *Annual Review of Neuroscience*, 16, 223–243.

Chait, M., Poeppel, D., de Cheveigne, A., & Simon, J. Z. (2007). Processing asymmetry of transitions between order and disorder in human auditory cortex. *Journal of Neuroscience*, 27(19), 5207–5214.

Chase, S. M., & Young, E. D. (2006). Spike-timing codes enhance the representation of multiple simultaneous sound-localization cues in the inferior colliculus. *Journal of Neuroscience*, 26, 3889–3898.

Chechik, G., Anderson, M. J., Bar-Yosef, O., Young, E. D., Tishby, N., & Nelken, I. (2006). I. Reduction of information redundancy in the ascending auditory pathway. *Neuron*, 51, 359–368.

Cherry, C. (1961). Two ears–but one world. In W. A. Rosenblith (Ed.). *Sensory communication* (pp. 99–117). New York: MIT Press/John Wiley & Sons.

Cherry, C. (1966). *On human communication*. Cambridge, MA: MIT Press.

de Cheveigné, A. (2005). Pitch perception models. In C. J. Plack, A. J. Oxenham, R. Fay & A. N. Popper (Eds.), *Pitch: Neural coding and perception* (pp. 169–233). New York: Springer.

Chow, K. L. (1951). Numerical estimates of the auditory central nervous system of the rhesus monkey. *Journal of Comparative Neurology*, 95, 159–175.

Clarke, S., & Thiran, A. B. (2004) Auditory neglect: What and where in auditory space. *Cortex*, 40(2), 291–300.

Damasio, A. R., & Damasio, H. (1994). Cortical systems for retrieval of concrete knowledge: The convergence zone framework. In C. Koch & J. L. Davis (Eds.), *Large-scale neuronal theories of the brain* (pp. 61–74), Cambridge, MA: MIT Press.

Delgutte, B., & Kiang, N. Y. S. (1984). Speech coding in the auditory nerve: I. Vowel-like sounds. *Journal of the Acoustical Society of America*, 75(3), 866–878.

Demany, L., & Semal, C. (2007). The role of memory in auditory perception. In W. A. Yost, A. N. Popper, & R. R. Fay (Eds.), *Auditory perception of sound sources* (pp. 77–113). New York: Springer.

DeWeese, M. R., Hromadka, T., & Zador, A. M. (2005). Reliability and representational bandwidth in the auditory cortex. *Neuron*, 48, 479–488.

Eggermont, J. J. (1990). *The correlative brain: Theory and experiment in neural interaction*. Berlin: Springer.

Emmers, R. (1981). *Pain: A spike-interval coded message in the brain*. New York: Raven Press.

Evans, E. F., Pratt, S. R., Spenner, H., & Cooper, N. P. (1992). Comparisons of physiological and behavioural properties: Auditory frequency selectivity. In Y. Cazals, K. Horner, & L. Demany (Eds.), *Auditory physiology and perception*, Vol. 83 (pp. 159–169). Oxford: Pergamon.

Fishman, Y. I., Reser, D. H., Arezzo, J. C., & Steinschneider, M. (2000). Complex tone processing in primary auditory cortex of the awake monkey. I. Neural ensemble correlates of roughness. *Journal of the Acoustical Society of America*, 108, 235–246.

Fritz, J., Shamma, S., Elhilali, M., & Klein, D. (2003). Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nature Neuroscience*, 6, 1216–1223.

Fritz, J., Mishkin, M., & Saunders, R. C. (2005). In search of an auditory engram. *Proceedings of the National Academy of Sciences of the USA*, 102, 9359–9364.

6–Furukawa, S., Xu, L., & Middlebrooks, J. C. (2000). Coding of sound-source location by ensembles of cortical neurons. *Journal of Neuroscience*, 20, 1216–1228.

Garde, M. M., & Cowey, A. (2000). "Deaf hearing": Unacknowledged detection of auditory stimuli in a patient with cerebral deafness. *Cortex*, 36(1), 71–79.

Greene, P. H. (1962). On looking for neural networks and "cell assemblies" that underlie behavior. I. Mathematical model. II. Neural realization of a mathematical model. *Bulletin of Mathematical Biophysics*, 24, 247–275, 395–411.

Handel, S. (1989). *Listening: An introduction to the perception of auditory events*. Cambridge, MA: MIT Press.

Heil, P. (1997). Auditory cortical onset responses revisited. I. First-spike timing. *Journal of Neurophysiology*, 77, 2616–2641.

Hromadka, T., DeWeese, M. R., & Zador, A. M. (2008). Sparse representation of sounds in the unanesthetized auditory cortex. *Public Library of Science (PLoS) Biology*, 6, e16.

Huetz, C., Gourevitch, B., & Edeline, J. M. (2011). Neural codes in the thalamocortical auditory system: From artificial stimuli to communication sounds. *Hearing Research*, 271(1–2), 147–158.

Huron, D. B. (2006). *Sweet anticipation: Music and the psychology of expectation*. Cambridge, MA: MIT Press.

Izhikevich, E. M. (2006). Polychronization: Computation with spikes. *Neural Computation*, 18, 245–282.

Jeffress, L. A. (1948). A place theory of sound localization. *Journal of Comparative and Physiological Psychology*, 41, 35–39.

Kayser, C., Petkov, C. I., & Logothetis, N. K. (2008). Visual modulation of neurons in auditory cortex. *Cerebral Cortex*, 18, 1560–1574.

Koch, C. (2004). *The quest for consciousness: A neurobiological approach*. Denver: Roberts & Co.

Köhler, W. (1947). *Gestalt psychology: An introduction to new concepts in modern psychology*. New York: Liveright.

Kowalski, N., Depireux, D. A., & Shamma, S. A. (1996). Analysis of dynamic spectra in ferret primary auditory cortex. I. Characteristics of single-unit responses to moving ripple spectra. *Journal of Neurophysiology*, 76, 3503–3523.

Lamme, V. A. (2006). Towards a true neural stance on consciousness. *Trends in Cognitive Sciences*, 10, 494–501.

Laurent, G. (2006). Shall we even understand the fly's brain? In J. L. van Hemmen & T. J. Sejnowski (Eds.), *23 problems in systems neuroscience* (pp. 3–21). Oxford: Oxford University Press.

Lee, K. M., Skoe, E., Kraus, N., & Ashley, R. (2009). Selective subcortical enhancement of musical intervals in musicians. *Journal of Neuroscience*, 29, 5832–5840.

Licklider, J. C. R. (1959). Three auditory theories. In S. Koch (Ed.), *Psychology: A study of a science. Study I. Conceptual and systematic*, Vol. I: *Sensory, perceptual, and physiological formulations* (pp. 41–144). New York: McGraw-Hill.

von der Malsberg, C. (1994). The correlation theory of brain function. In E. Domany, J. L. van Hemmen, & K. Schulten (Eds.), *Models of neural networks II: Temporal aspects of coding and information processing in biological systems* (pp. 95–120). New York: Springer.

Marr, D. (1982). *Vision: A computational approach*. San Francisco: Freeman & Co.

McAdams, S., & Giordano, B. L. (2009). The perception of musical timbre. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford handbook of music psychology* (pp. 72–80). Oxford: Oxford University Press.

McCulloch, W. S. (1951). Why the mind is in the head. In L. A. Jeffress (Ed.), *Cerebral mechanisms of behavior* (pp. 42–111). New York: John Wiley & Sons.

McDermott, J., & Oxenham, A. (2008) Music perception, pitch, and the auditory system. *Current Opinion in Neurobiology*, 18, 452–463.

Mesulam, M. M. (2000). *Principles of behavioral and cognitive neurology*. New York: Oxford University Press.

Miller, L. M., Escabi, M. A., Read, H. L., & Schreiner, C. E. (2001). Functional convergence of response properties in the auditory thalamocortical system. *Neuron*, 32, 151–160.

Miller, L. M., Escabí, M. A., Read, H. L., & Schreiner, C. E. (2002). Spectrotemporal receptive fields in the lemniscal auditory thalamus and cortex. *Journal of Neurophysiology*, 87, 516–527.

Monahan, P. J., & Idsardi, W. J. (2010). Auditory sensitivity to formant ratios: Toward an account of vowel normalization. *Language and Cognitive Processes*, 25(6), 808–839.

Moore, B. C. (1973). Frequency difference limens for short-duration tones. *Journal of the Acoustical Society of America*, 54, 610–619.
Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, 118, 2544–2590.
Nelken, I., Fishbach, A., Las, L., Ulanovsky, N., & Farkas, D. (2003). Primary auditory cortex of cats: Feature detection or something else? *Biological Cybernetics*, 89, 397–406.
Palmer, A. R. (1992). Segregation of the responses to paired vowels in the auditory nerve of the guinea pig using autocorrelation. In M. E. H. Schouten (Ed.), *The auditory processing of speech* (pp. 115–124). Berlin: Mouton de Gruyter.
Panzeri, S., Brunel, N., Logothetis, N. K., & Kayser, C. (2009). Sensory neural codes using multiplexed temporal scales. *Trends in Neurosciences*, 33(3), 111–120.
Patel, A. D. (2008). *Music, language and the brain*. Oxford: Oxford University Press.
Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., & Griffiths, T. D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron*, 36, 767–776.
Penagos, H., Melcher, J. R., & Oxenham, A. J. (2004). A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. *Journal of Neuroscience*, 24, 6810–6815.
Perkell, D. H., & Bullock T. H. (1968). Neural coding. *Neurosciences Research Program Bulletin*, 6, 221–348.
Phillips, D. P., Semple, M. N., Calford, M. B., & Kitzes, L. M. (1994). Level-dependent representation of stimulus frequency in cat primary auditory cortex. *Experimental Brain Research*, 102, 210–226.
Phillips, D. P., Hall, S. E., & Boehnke, S. E. (2002). Central auditory onset responses, and temporal asymmetries in auditory perception. *Hearing Research*, 167, 192–205.
Reprinted in W. S. McCulloch (Ed.), *Embodiments of mind* (pp. 46–66). Cambridge, MA: MIT Press, 1965.Pitts, W., & McCulloch, W. S. (1947). How we know universals: The perception of auditory and visual forms. *Bulletin of Mathematical Biophysics*, 9, 127–147. Reprinted in W. S. McCulloch (Ed.), *Embodiments of mind* (pp. 46–66). Cambridge, MA: MIT Press, 1965.
Pollen, D. A. (2008). Fundamental requirements for primary visual perception. *Cerebral Cortex*, 18, 1991–1998.
Rauschecker, J. P., & Tian, B. (2004). Processing of band-passed noise in the lateral auditory belt cortex of the rhesus monkey. *Journal of Neurophysiology*, 91, 2578–2589.
Rieke, F., Warland, D., de Ruyter van Steveninck, R., & Bialek, W. (1997). *Spikes: Exploring the neural code*. Cambridge, MA: MIT Press.
Sadagopan, S., & Wang, X. (2008). Level invariant representation of sounds by populations of neurons in primary auditory cortex. *Journal of Neuroscience*, 28(13), 3415–3426.
Schreiner, C. E., & Langner, G. (1988). Coding of temporal patterns in the central auditory nervous system. In G. Edelman (Ed.), *Auditory function: Neurobiological bases of hearing* (pp. 337–361). New York: John Wiley & Sons.
Schroeder, C. E., & Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neurosciences*, 32(1), 9–17.
Shamma, S. A., & Micheyl, C. (2010). Behind the scenes of auditory perception. *Current Opinion in Neurobiology*, 20, 361–366.
Shamma, S. A., Elhilali, M., & Micheyl, C. (2010). Temporal coherence and attention in auditory scene analysis. *Trends in Neurosciences*, 34(3), 114–123.
1–Singer, W. (2003). Synchronization, binding, and expectancy. In M. A. Arbib (Ed.). *The handbook of brain theory and neural networks*, 2nd ed. (pp. 1136–1143). Cambridge, MA: MIT Press.
Snyder, B. (2000). *Music and memory*. Cambridge, MA: MIT Press.
Snyder, B. (2009). Memory for music. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford handbook of music psychology* (pp. 171–183). Oxford: Oxford University Press.
Thatcher, R. W., & John, E. R. (1977). *Foundations of cognitive processes*. Hillsdale, NJ: Lawrence Erlbaum.
Tian, B., & Rauschecker, J. P. (1994). Processing of frequency-modulated sounds in the cat's anterior auditory field. *Journal of Neurophysiology*, 71, 1959–1975.

Tononi, G., & Koch, C. (2008). The neural correlates of consciousness: An update. *Annals of the New York Academy of Sciences*, 1124, 239–61.

Trainor, L. J., & Zatorre, R. J. (2009). The neurobiological basis of musical expectations. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford handbook of music psychology* (pp. 171–183). Oxford: Oxford University Press.

Tramo, M. J., Cariani, P. A., Koh, C. K., Makris, N., & Braida, L. D. (2005). Neurophysiology and neuroanatomy of pitch perception: Auditory cortex. *Annals of the New York Academy of Sciences*, 1060, 148–174.

Tzounopoulos, T., Kim, Y., Oertel, D., & Trussell, L. O. (2004). Cell-specific, spike timing-dependent plasticities in the dorsal cochlear nucleus. *Nature Neuroscience*, 7, 719–725.

Ulanovsky, N., Las, L., Farkas, D., & Nelken, I. (2004). Multiple time scales of adaptation in auditory cortex neurons. *Journal of Neuroscience*, 24, 10440–10453.

Van Rullen, R., & Thorpe, S. J. (2001). Rate coding versus temporal order coding: What the retinal ganglion cells tell the visual cortex. *Neural Computation*, 13, 1255–1283.

Villa, A. E. (2000). Empirical evidence about temporal structure in multi-unit recordings. In R. Miller (Ed.), *Time and the brain* (pp. 1–52). Amsterdam: Harwood.

Wallace, M. N., Shackleton, T. M., & Palmer, A. R. (2002). Phase-locked responses to pure tones in the primary auditory cortex. *Hearing Research*, 172, 160–171.

Wang, D. (2002). *The time dimension for neural computation* (pp. 1–40). Columbus, OH: Center for Cognitive Science and the Department of Computer & Information Science, The Ohio State University.

Wang, X. (2007). Neural coding strategies in auditory cortex. *Hearing Research*, 229, 81–93.

Warren, J. D., Uppenkamp, S., Patterson, R. D., & Griffiths, T. D. (2003). Separating pitch chroma and pitch height in the human brain. *Proceedings of the National Academy of Sciences of the USA*, 100, 10038–10042.

Winer, J. A. (1992). The functional architecture of the medial geniculate body and the primary auditory cortex. In D. B. Webster, A. N. Popper, & R. R. Fay (Eds.), *The mammalian auditory pathway: Neuroanatomy* (pp. 222–286). New York: Springer.

Winkler, I., Denham, S. L., & Nelken, I. (2009). Modeling the auditory scene: Predictive regularity representations and perceptual objects. *Trends in Cognitive Sciences*, 13, 532–540.

Wiskott, L. (2006). How does our visual system achieve shift invariance? In J. L. van Hemmen (Ed.), *23 problems in systems neuroscience* (pp. 322–340). Oxford: Oxford University Press.