

Neural representation of pitch through temporal autocorrelation

Peter Cariani^{1,2}, Mark Tramo^{1,3}, and Bertrand Delgutte^{1,2}

¹Eaton Peabody Laboratory, Massachusetts Eye & Ear Infirmary
²Department of Otology and Laryngology, Harvard Medical School
³Department 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.

Six stimuli that produce a low pitch at 160 Hz

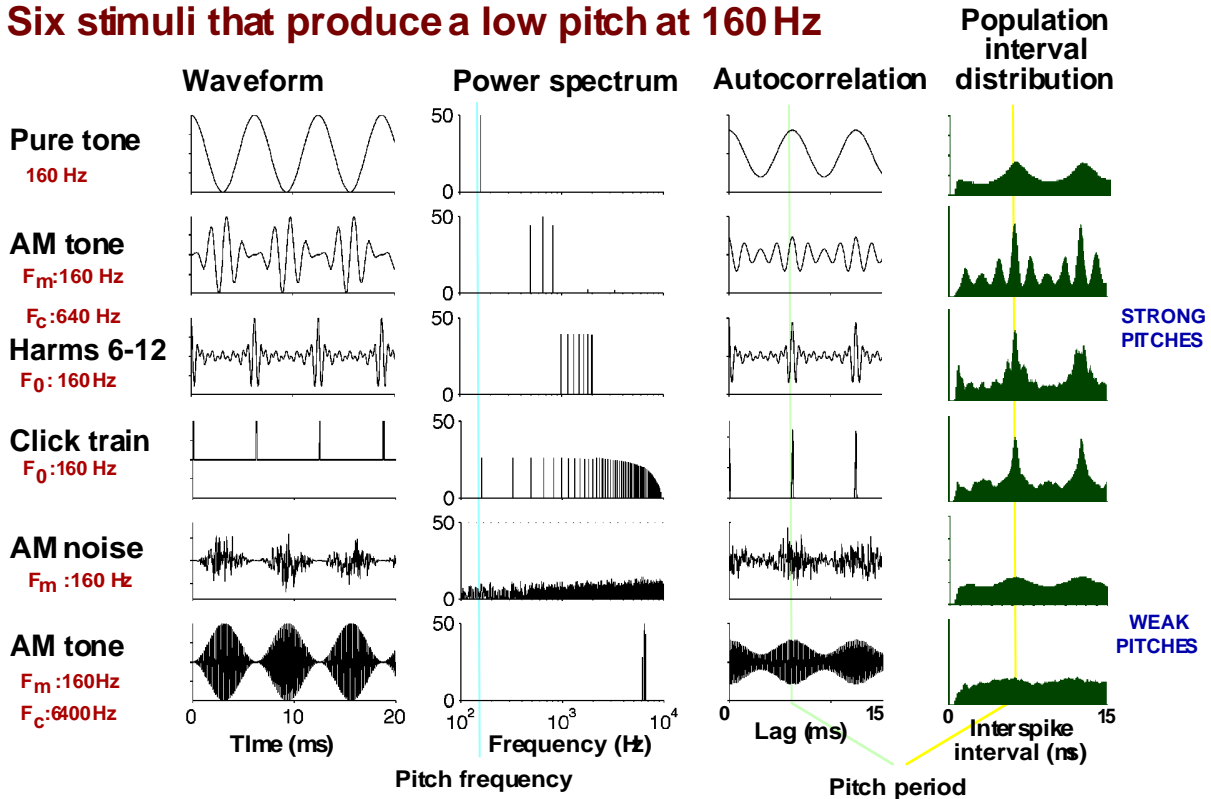


Figure 1. Neural responses to four stimuli evoking a pitch at 160 Hz but differing in pitch salience. Left to right: Stimulus waveform, power spectrum, short term autocorrelation function, and population-interval distribution for each stimulus. Population-interval distributions are constructed by summing 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 ($F_c=640$ Hz, $F_m=160$ Hz). C. Harmonic complex (harmonics 6-12, of 160 Hz). D. Unipolar click train ($F_0=160$ Hz.) E. AM tone with a high frequency carrier ($F_c=6400$ Hz, $F_m=160$ Hz). F. AM broadband noise with $F_m=160$ 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].

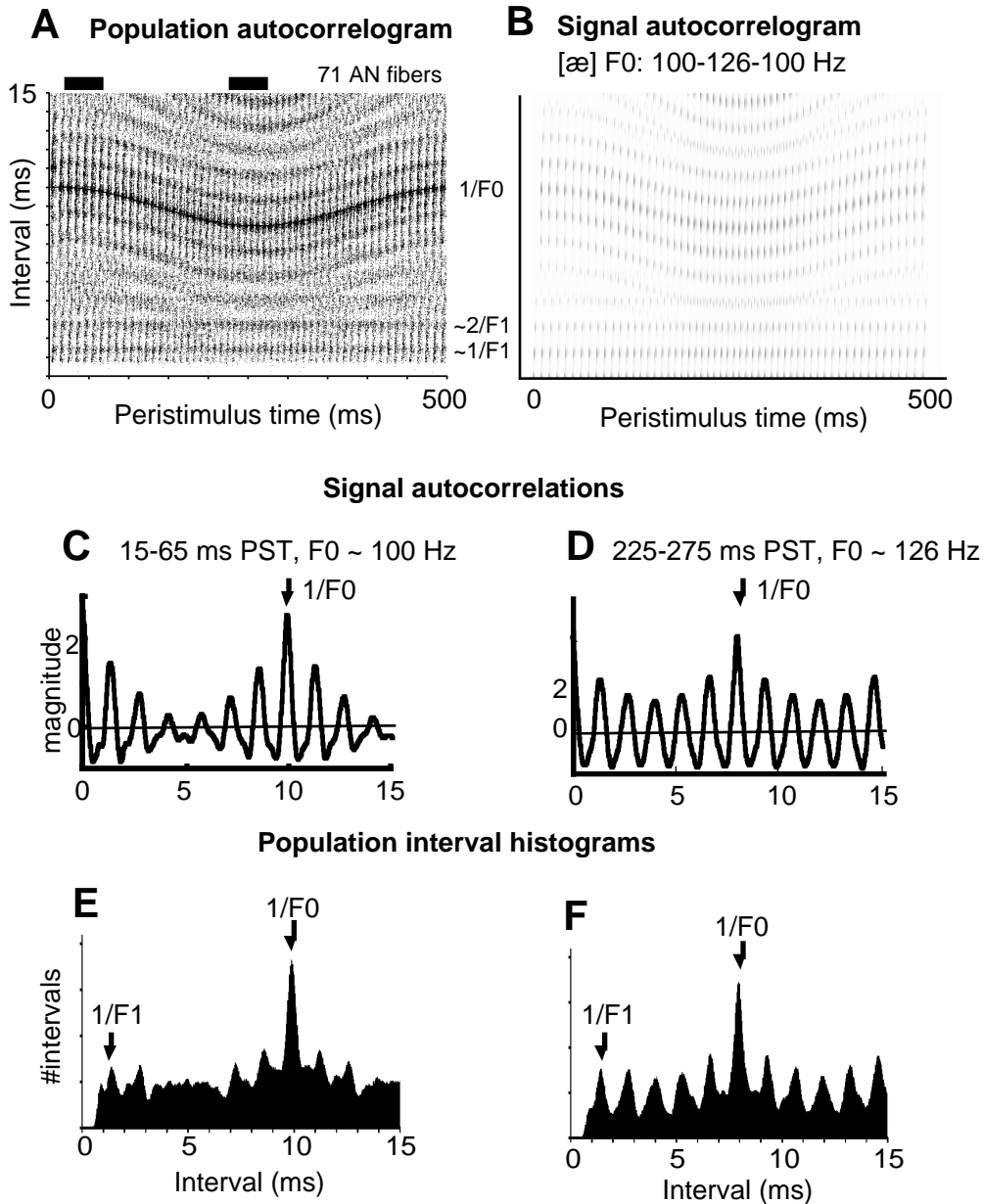


Figure 2. Autocorrelation-like representations of the vowel [æ]. Stimulus: Five-formant synthetic vowel [æ], as in "had". Formants are at 750, 1450, 2450, 3350, 3850 Hz, and the fundamental frequency F_0 (heard as the voice pitch) sinusoidally varies between 100-126 Hz. A. Neural response. Running population-interval 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 occurrence 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, $1/F_0$, 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 $1/F_0$ 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.

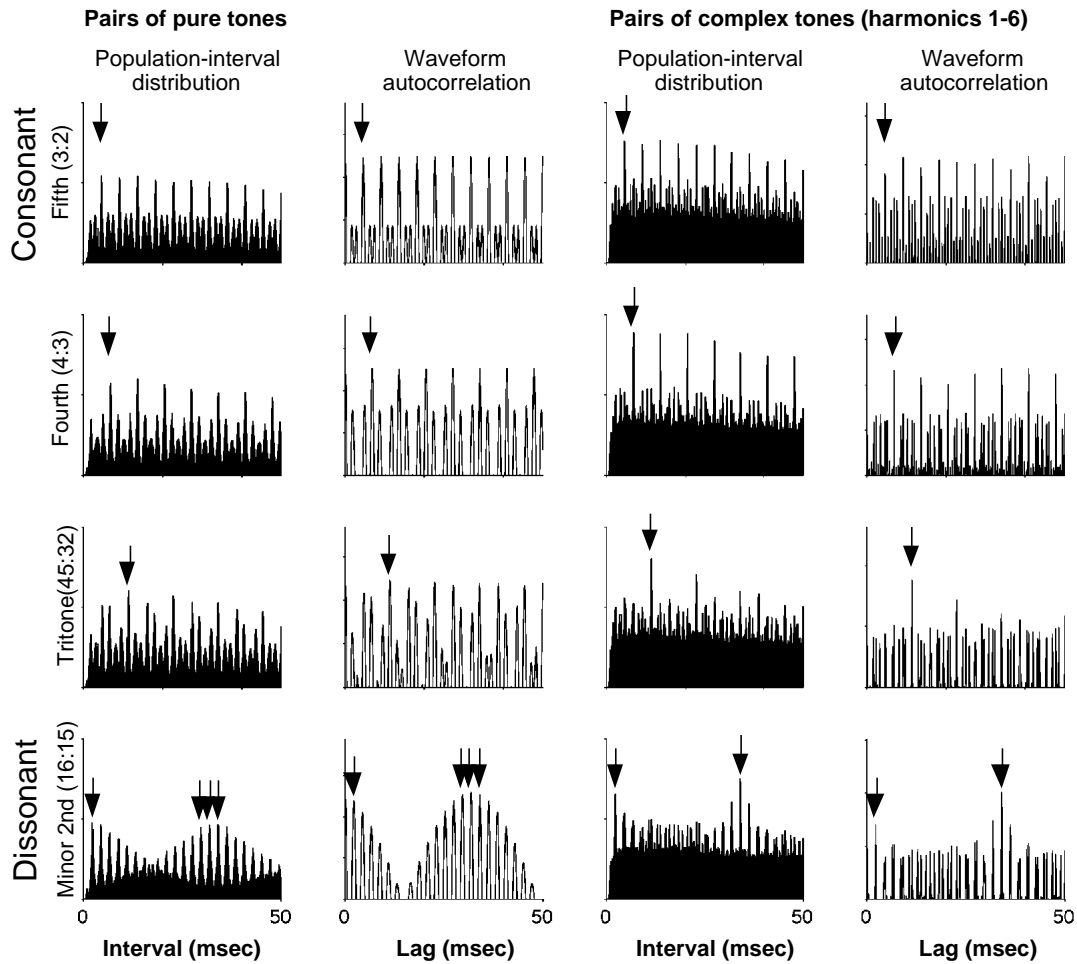


Figure 3. Population-interval neural representations and corresponding stimulus autocorrelation functions for five musical intervals. Stimuli were pairs of pure tones or pairs of complex tones consisting of harmonics 1-6 (equal amplitudes), with the lower fundamental frequency = 440 Hz. For complex tones, 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 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: $f(\tau) = \sum S(t) S(t-\tau)$, 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 $F_0 = 100$ Hz and when $F_0 = 126$ 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 $2F_0$).

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 F_0 -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. Sruлович, 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.