

Temporal codes, timing nets, and music perception

Peter Cariani

Eaton Peabody Laboratory, Massachusetts Eye and Ear Infirmary
243 Charles St., Boston, MA 02114, USA

Department of Otology and Laryngology
Harvard Medical School

email: cariani@mac.com, www.cariani.com
fax: 413-618-4025

Revised version, 6/20/01

For Journal of New Music Perception, vol. 30, issue 2 (2001)
Special issue on Rhythm Perception, Periodicity, and Timing Nets

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.

BIOGRAPHICAL SKETCH

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 and collaborating with Dr. Mark Tramo in the investigation of the neural representation of pitch in the auditory cortex.

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 subservise 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, and these 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 subservise 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 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 that 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 dynamic range) 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: repeating temporal patterns related by simple ratios contain common subpatterns that potentially explain octave-equivalences and other Pythagorean observations. But it is another step entirely to hypothesize that the brain itself uses a time code that replicates the temporal structure of sounds, and 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-). 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 decision-theoretic models of pure tone pitch discrimination (Delgutte, 1996; Goldstein & Sruлович, 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 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; Cariani, 2001, in press). 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 to arrive at 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 are 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. This 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 (F_0) 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 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 (Figure 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 through 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 (Figure 2) can be directly compared with the observed neural responses to the same single-formant vowel (Figure 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 (Figure 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 (that are amplitude-symmetric with zero-mean), 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 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 work 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 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, & Braid, 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 $1/F_0$ 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 (an upward dc shift). 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 $F_0 = 100$ (1:1), 200 (2:1), and 300 (3:1) share intervals at 10 and 20 ms. Distributions for $F_0 = 100$ 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 $F_0 = 200$ 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 sinusoidal 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)(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 τ s less than or equal

to 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 the population-interval distributions of complex tones show additional correlation peaks has to do with correlations produced by 1) spectral overlap, i.e. by harmonics that are common to the two notes and 2) by octave-relations, i.e. harmonics related by octaves (2:1) and twelfths (3:1), i.e. the positive correlations present between the pure tones.

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 through 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.

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). Comparable results were also obtained 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 no apparent representation in auditory frequency maps, clearly subharmonics are present 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 possess 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 character, 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 kinds 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, and 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 thus 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 & Sruлович, 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 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, respectively, connectionist networks, time-delay neural networks, and neural timing nets.

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

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 not significant for the encoding of relevant distinctions. 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 coincidence 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; Cariani, 2001, in press). 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 (Figure 7A) act as temporal pattern sieves to extract common periodicities in their inputs, and thus are relevant to perceptual comparisons of pitch, timbre, and rhythm. Recurrent timing nets (Figure 7B) build up temporal patterns that recur in their inputs to form temporal expectations of what is to follow. We will discuss how recurrent timing networks may be applied to the formation of rhythmic expectations.

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. For the most part, while conceptions incorporate notions of reverberating circuits (Hebb, 1949), it has been spatial patterns (Grossberg, 1988) and sequences of neural activations (e.g. (McCulloch, 1969)), and not their time structure, that has been utmost in their minds. 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 (Figure 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 similarly have been present in the two inputs with the leading spikes of each interval obeying the temporal contiguity constraint. 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 $A_k(\tau) = \sum [S_i(t)S_j(t-D_k)][S_i(t)S_j(t-D_k)-\tau]$, in which the product of the output spike train and itself delayed is summed together for each all-order interval of duration τ . 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. $SAC(\tau) = \sum [A_i]$. 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 (Figure 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 τ , and the other has N such intervals, $M \cdot N$ τ -length intervals will appear in the outputs (Figure 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 (Figure 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 (Figures 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.

Table II. 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

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 fundamental 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 II. 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 III. The “pipe organ” and “alto sax” voices have very similar short-interval patterns, but these differ substantially from that of the “piano” voice.

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; Cariani, 2001, in press).

	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

Table III. Correlations for 0-2 ms (bottom) and 2-20 ms intervals (top)

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 Figures 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; Cariani, 2001, in press). 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 are embodied 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 pattern has properties, a pattern of major and minor peaks, that reflect combinations of partials. Thus the forms of both the whole and of the parts 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 (the common periodicity) and discrete qualities (presence or absence of a common periodicity).

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 that terminate on the element that generated them. 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, “re-entrant” connections between cortical areas (McCulloch, 1947). The impressive array of recurrent 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 an 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 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 $A_k(t) = \min(S_{\text{direct}}(t), B * S_{\text{direct}}(t) * S_{\text{loop}}(t))$, where $A_k(t)$ is the output of coincidence element k associated with delay loop of recurrence time LD_k , $S_{\text{direct}}(t)$ is the incoming direct input signal, and $S_{\text{loop}}(t)$ is the incoming signal circulating in the loop.

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 sub-meters 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 (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 11. The running rms of the waveform of the audio recording of the piano performance 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 (second panel) was analyzed using running autocorrelation and a recurrent timing net.

The running autocorrelation (RAC) of the Ligeti fragment is shown in the third panel. 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 (τ) 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 autocorrelation can be compared with other running displays of time structure: cochleograms (Lyon & Shamma, 1996; Slaney & Lyon, 1993), tempograms and the strobed auditory images of (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 Figure 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 the *form* of the periodic pattern which retains 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, 2001, in press). 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.

ACKNOWLEDGEMENTS

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.

Tables

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

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 2. Correlations between population interval distributions for waveforms A-D

	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

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

FIGURE CAPTIONS

Figure 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 ($F_0=80$ Hz, pitch period (double arrow) $1/F_0=12.5$ 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).

Figure 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.

Figure 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 responses 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.

Figure 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.

Figure 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.

Figure 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.

Figure 7. Neural timing nets. Top. 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. Bottom: 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.

Figure 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.

Figure 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.

Figure 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.

Figure 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.

Figure 12. Analysis of Ligeti fragment by autocorrelograms 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.

Embedded fonts: Times, **times-bold**, helvetica, **helvetica-bold**, symbol, **symbol-bold**

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.
- Boomsliker, 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. *J. Neurophysiol.*, 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(4), 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(1-3), 59-83.
- Cariani, P. (2001c). Temporal coding of sensory information in the brain. *Acoust. Sci. & Tech.*, 22(2), 77-84.
- Cariani, P. (2001, in press). Neural timing nets. *Neural Networks*, 16.
- 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. *J. Neurophysiol.*, 76(3), 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. *J. Neurophysiol.*, 76(3), 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. *J. Acoust. Soc. Am.*, 77(3), 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(2), 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. *J. Acoust. Soc. Am.*, 103(3), 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. *J. Acoust. Soc. Am.*, 83(2), 687-695.
- Demany, L., & Semal, C. (1990). Harmonic and melodic octave templates. *J. Acoust. Soc. Am.*, 88(5), 2126-2135.
- DeWitt, L. A., & Crowder, R. G. (1987). Tonal fusion of consonant musical intervals: The oomph in Stumpf. *Perception and Psychophysics*, 41(1), 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), 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. *J. Acoust. Soc. Am.*, 54(6), 1496-1516.
- Goldstein, J. L., & Sruлович, 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. *J. Acoust. Soc. Am.*, 69(2), 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(5), 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. *J. Acoust. Soc. Am.*, *104*(4), 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 interdisciplinary discussion. II. *Naturwissenschaften*, *79*(8), 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. *Exp. Brain Res.*, *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*(4), 481-510.
- Leman, M., & Carreras, F. (1997). 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.
- 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. *J. Acoust. Soc. Am.*, *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. *J. Acoust. Soc. Am.*, 106(5), 2679-2692.
- Meddis, R., & Hewitt, M. J. (1991a). Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I. Pitch identification. *J. Acoust. Soc. Am.*, 89(6), 2866-2882.
- Meddis, R., & Hewitt, M. J. (1991b). Virtual pitch and phase sensitivity of a computer model of the auditory periphery. II. Phase sensitivity. *J. Acoust. Soc. Am.*, 89(6), 2883-2894.
- Meddis, R., & O'Mard, L. (1997). A unitary model of pitch perception. *J. Acoust. Soc. Am.*, 102(3), 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, 4th Ed.* (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. *J. Acoust. Soc. Am.*, 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. *J. Acoust. Soc. Am.*, 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. *J. Acoust. Soc. Am.*, 98(4), 1890-1894.
- Perkell, D. H., & Bullock, T. H. (1968). Neural Coding. *Neurosciences Research Program Bulletin*, 6(3), 221-348.
- Pressnitzer, D., Patterson, R. D., & Krumboltz, K. (2001). The lower limit of melodic pitch. *J. Acoust. Soc. Am.*, 109(5), 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. *J. Acoust. Soc. Am.*, 103(1), 588-601.
- Scheirer, E. D. (2000). *Music Listening Systems*. Cambridge, Mass.: Ph.D. Dissertation, M.I.T.
- Schneider, A. (1997). "Verschmelzung", tonal fusion, and consonance: 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". *J. New Music Research*.
- Schouten, J. F. (1940). The residue, a new concept in subjective sound. *Proc. Koninkl. Ned. Akad. Wetenschap.*, 43, 356-365.
- Schouten, J. F., Ritsma, R. J., & Cardozo, B. L. (1962). Pitch of the residue. *J. Acoust. Soc. Am.*, 34(8 (Part 2)), 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.
- Shamma, S., & Sutton, D. (2000). The case of the missing pitch templates: How harmonic templates emerge in the early auditory system. *J. Acoust. Soc. Am.*, 107(5), 2631-2644.
- Shepard, R. N. (1964). Circularity in judgments of relative pitch. *J. Acoust. Soc. Am.*, 36(12), 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. *J. Acoust. Soc. Am.*, 55(5), 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. , 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. *J. Acoust. Soc. Am.*
- 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. *J. Acoust. Soc. Am.*, 103(4), 2051-2059.
- Wiegand, L. (2001). Searching for the time constant of neural pitch extraction. *J. Acoust. Soc. Am.*, 109(3), 1082-1091.

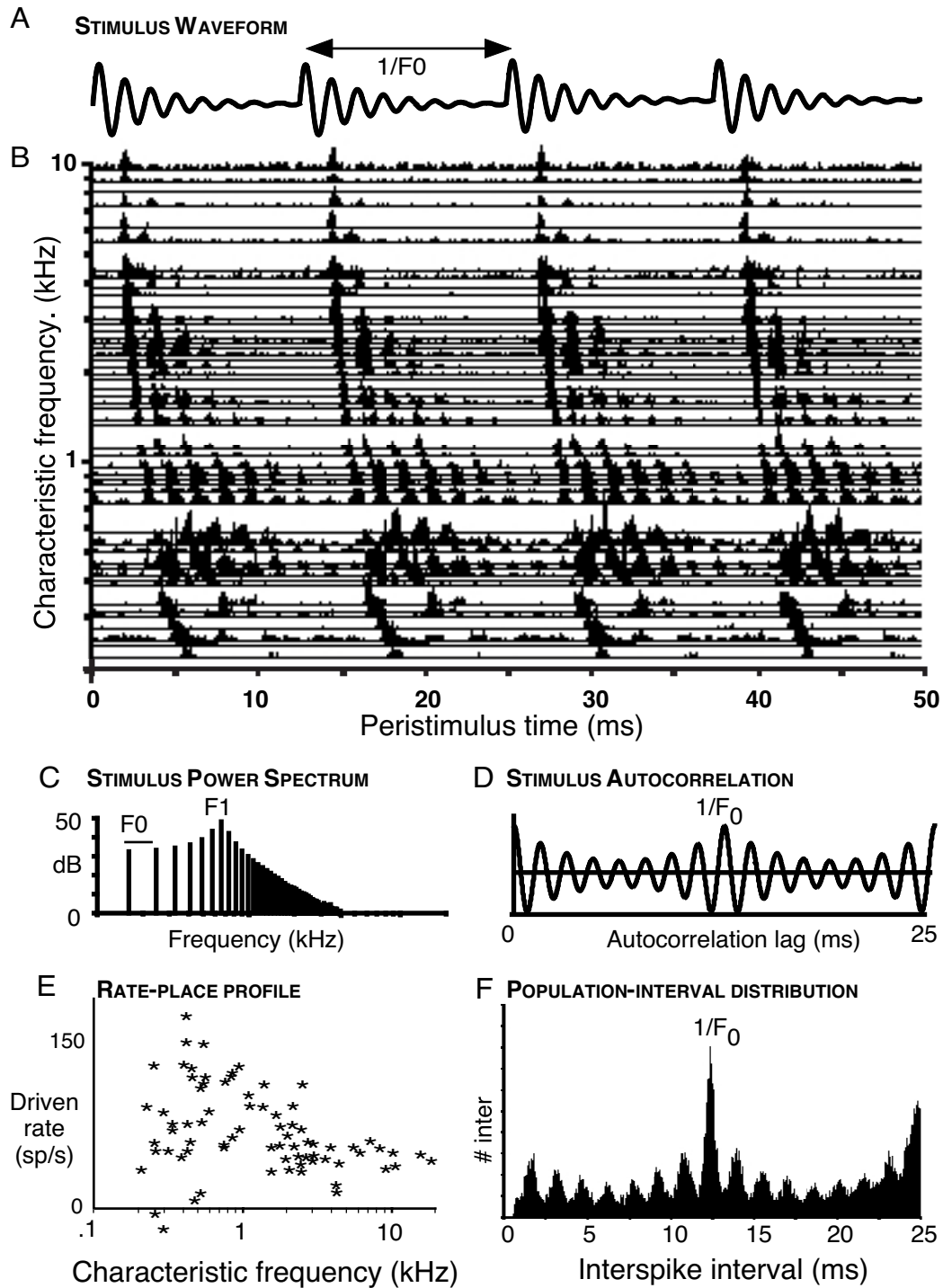


Figure 1. Temporal coding of musical pitch in the auditory nerve. Auditory nerve responses to a harmonic complex tone with a single formant. A. Stimulus waveform. A strong, low voice pitch is heard at the fundamental ($F_0=80$ Hz, pitch period (double arrow) $1/F_0=12.5$ ms). B. Peristimulus time histograms 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).

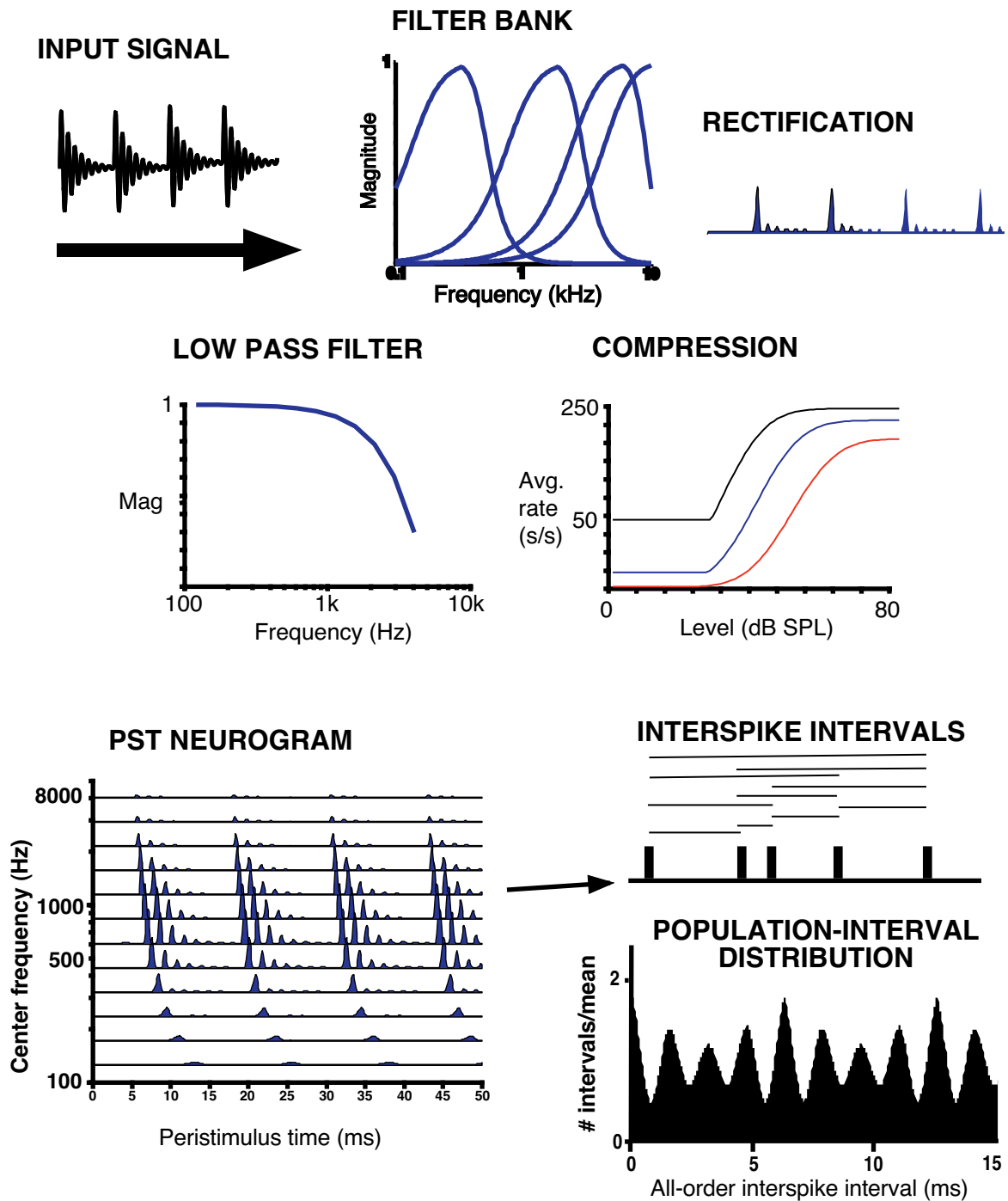


Figure 2. Auditory nerve array simulation for the estimation of population-interval distributions. An input signal is passed through a bank of bandpass filters, half-wave rectified, low pass filtered, and compressed using three rate-level functions to produce post-stimulus time (PST) histograms for each 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.

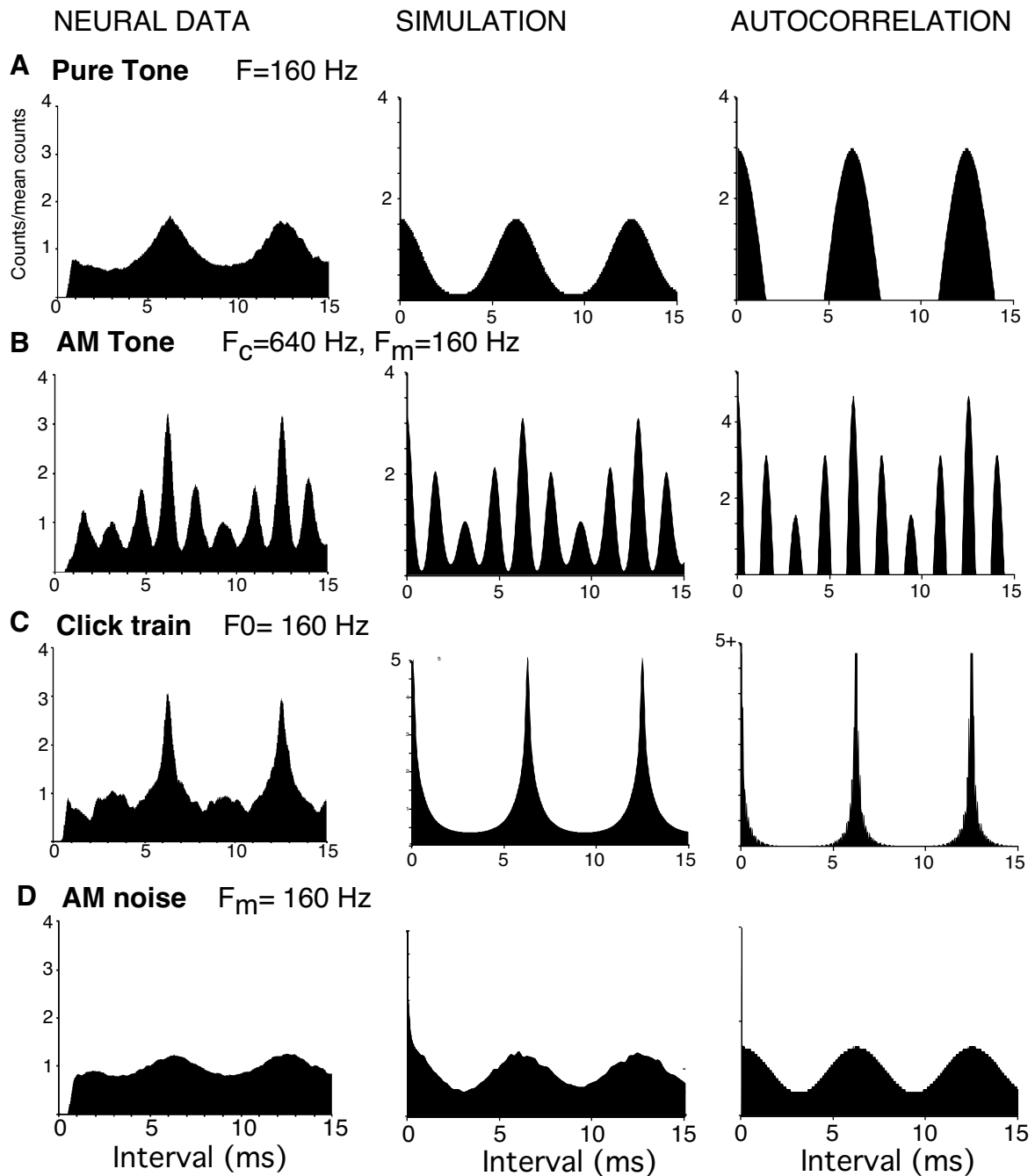


Figure 3. Comparisons of population-interval distributions and autocorrelation function for six stimuli that produce a low pitch at 160 Hz. Left. Population interval distributions estimated from recorded responses of 50-100 auditory nerve fibers in Dial-anesthetized cats (Cariani & Delgutte, 1996). Middle.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

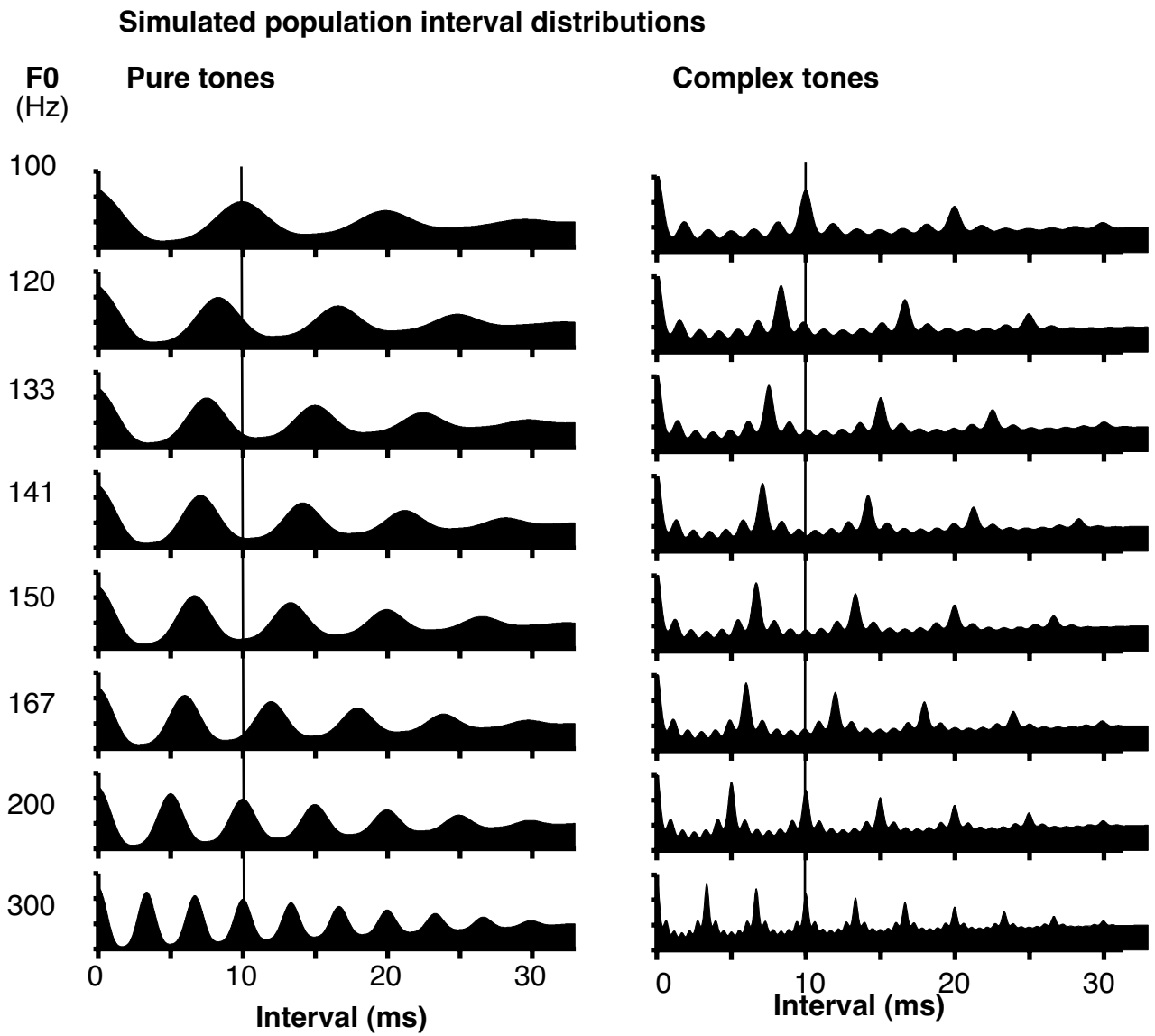


Figure 4. Similarities between population-interval representations associated with different fundamental frequencies. Simulated population-interval distributions for pure tones (left) and complex tones (right) consisting of harmonics 1-6.

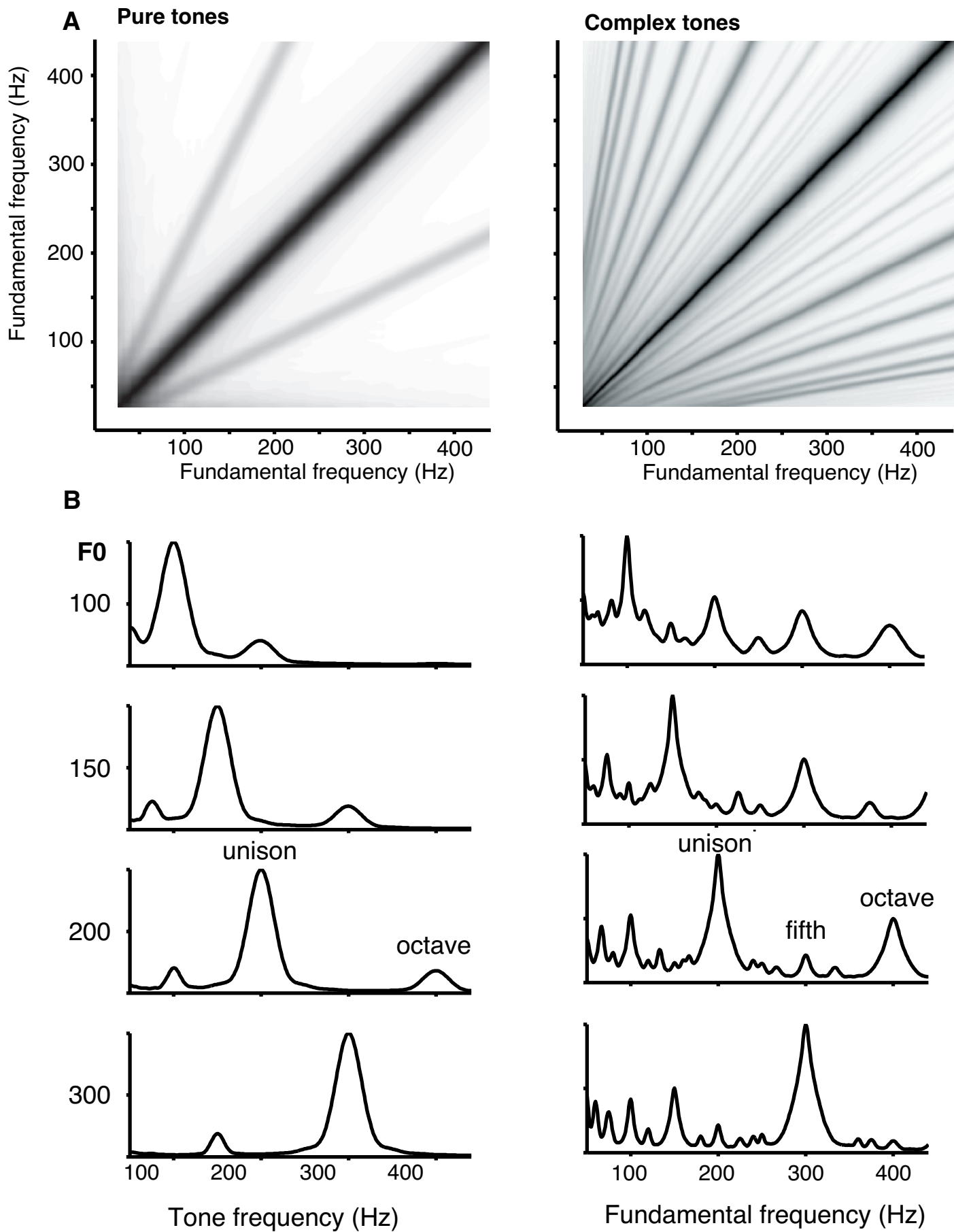


Figure 5. Tonal structure and pattern similarities between population-interval distributions. Top. Map of correlation coefficients between all pairs of simulated population-interval distributions produced by pure and complex tones with fundamentals ranging from 1-440 Hz. B. Cross-sectional correlation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

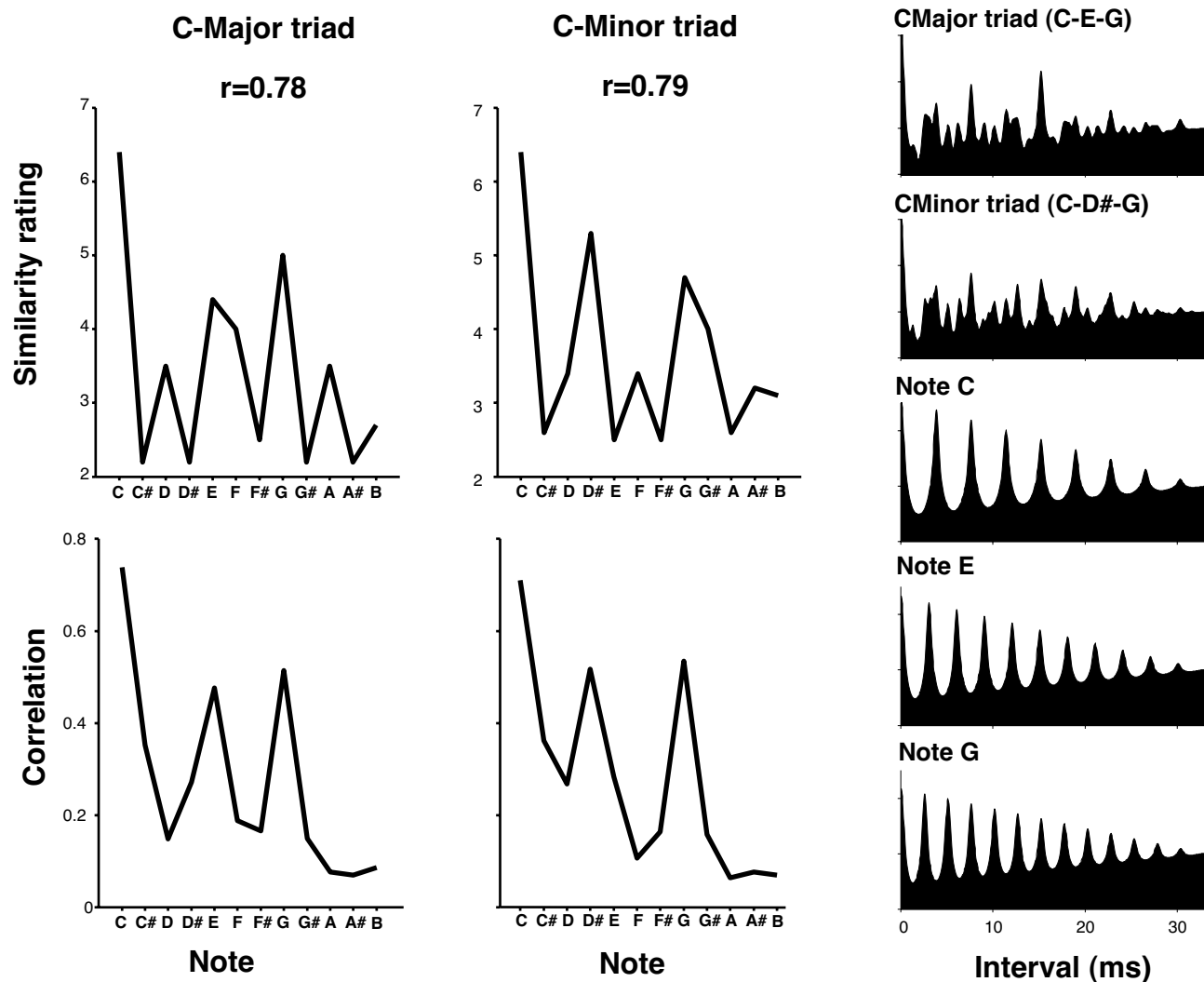
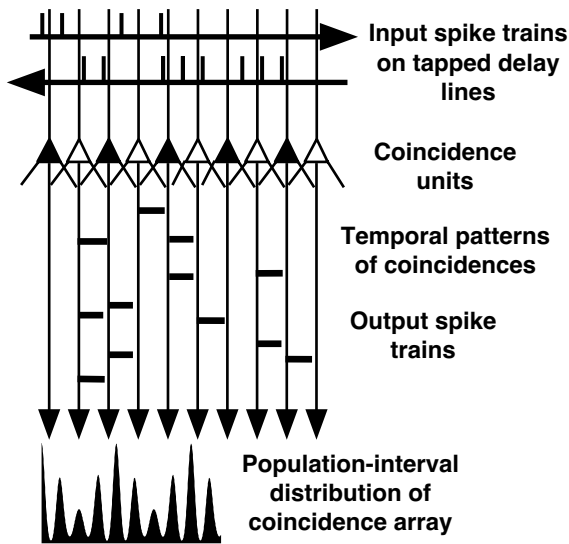


Figure 6. Comparison of interval-based measures of note-chord similarity with human judgments. Left top. Results of probe tone experiments: human ratings of how well a note fits in with a preceding chord (Krumhansl, 1990). Chords were either C-major (CEG) or C-minor (CD#G) note triads. Notes consisted 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.

A FEEDFORWARD TIMING NET



B RECURRENT TIMING NET

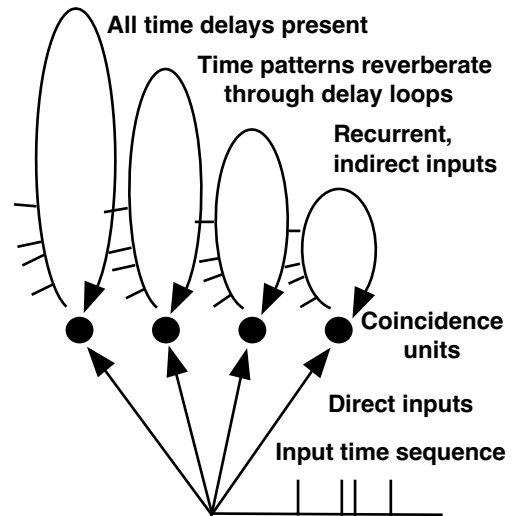


Figure 7. Neural timing nets. Top. A simple feedforward timing net consisting of two tapped delay lines and a linear array of coincidence detectors. Outputs of coincidence detectors contain only temporal patterns that are common to the two inputs. The population interspike interval distribution of the outputs of the net reflects a comparison between the interval distributions of the two inputs. Bottom: 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.

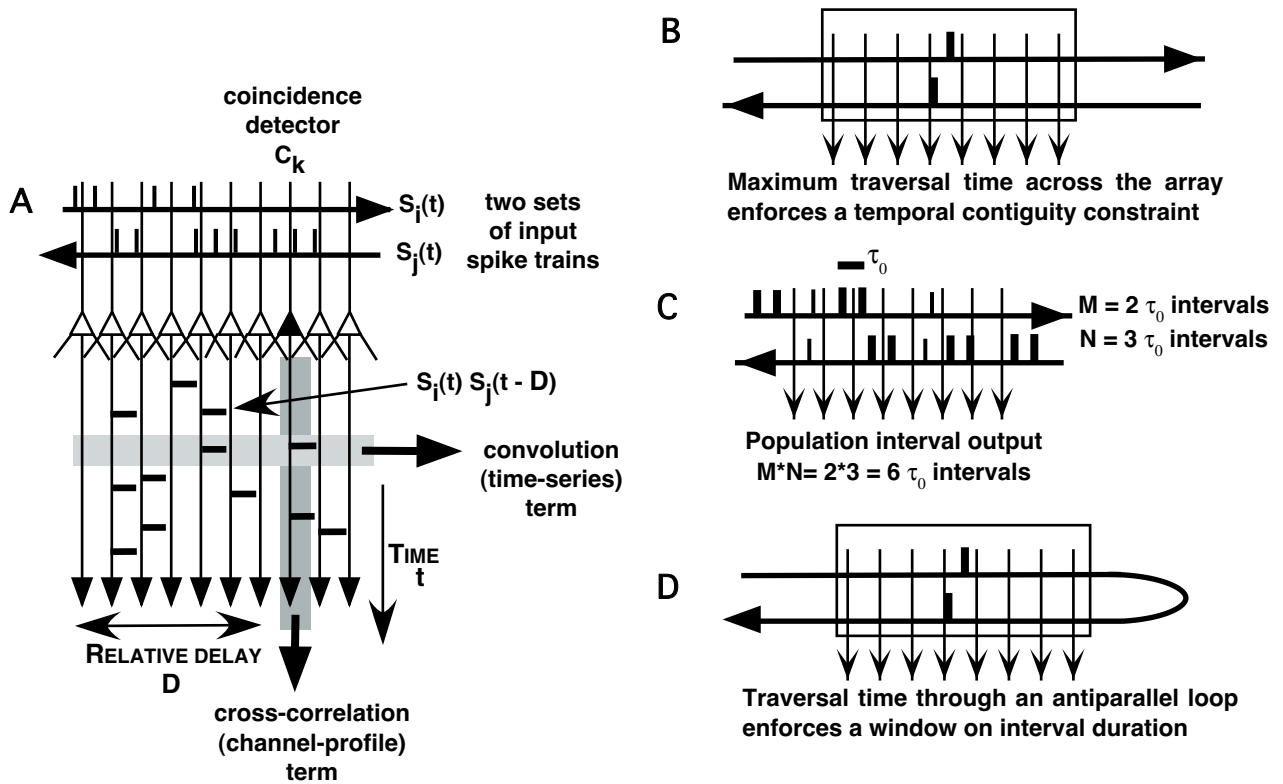


Figure 8. A simple feedforward timing net. A. General schematic of a coincidence array traversed by tapped delay lines. Summation over time in each output channel yields the cross-correlation function, while summation over output channels for each time yields the convolution of the two inputs. B. The 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.

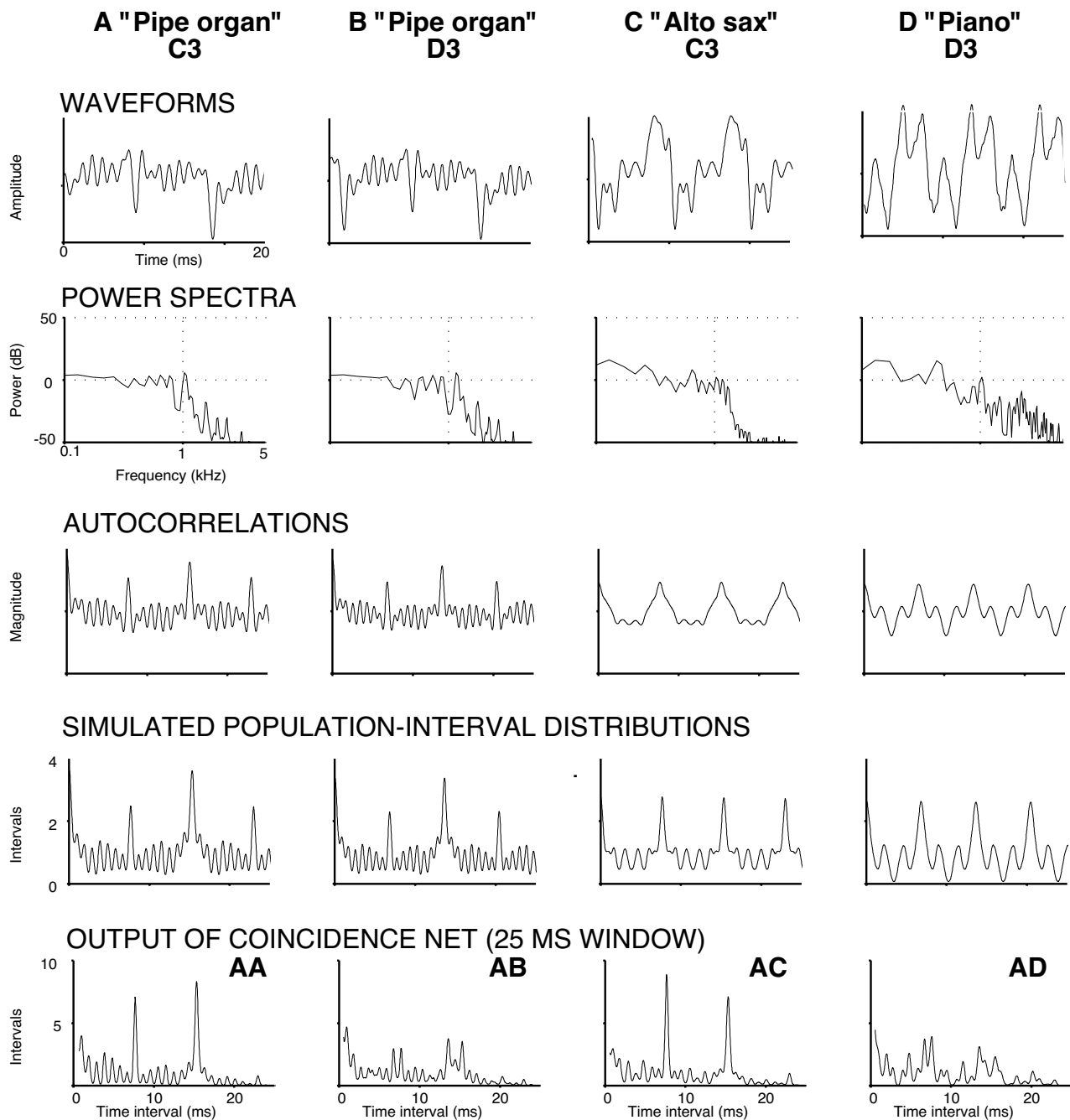
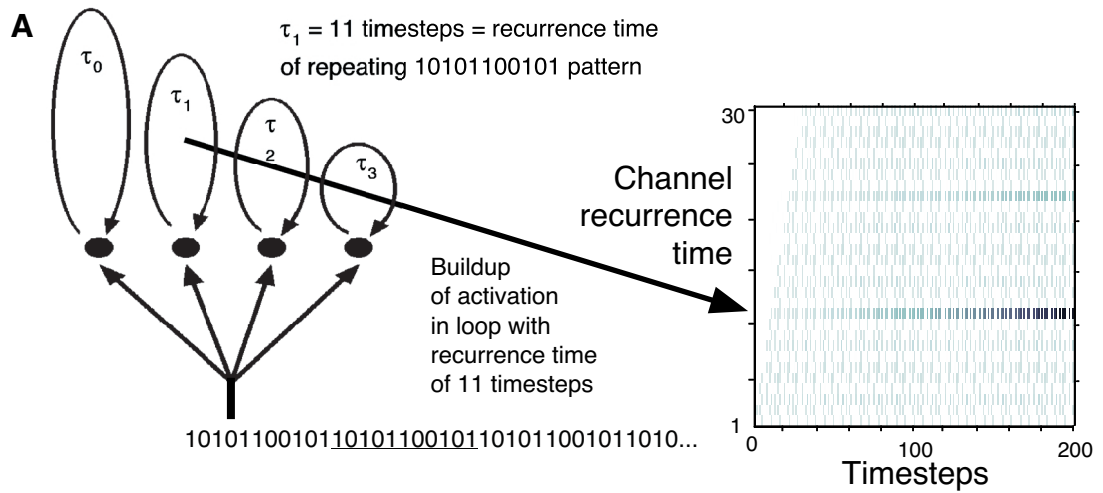


Figure 9. Pitch matching irrespective of timbral difference. Waveforms, power spectra, autocorrelations, and simulated population-interval distributions are shown for four synthesized waveforms (Yamaha PSR-76). Complex tones A and B have similar spectral envelopes and produce the same timbre. Tones A 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.



B La Marseillaise rhythm
 110011000100010001000100000011001100110001000000010011000000000...

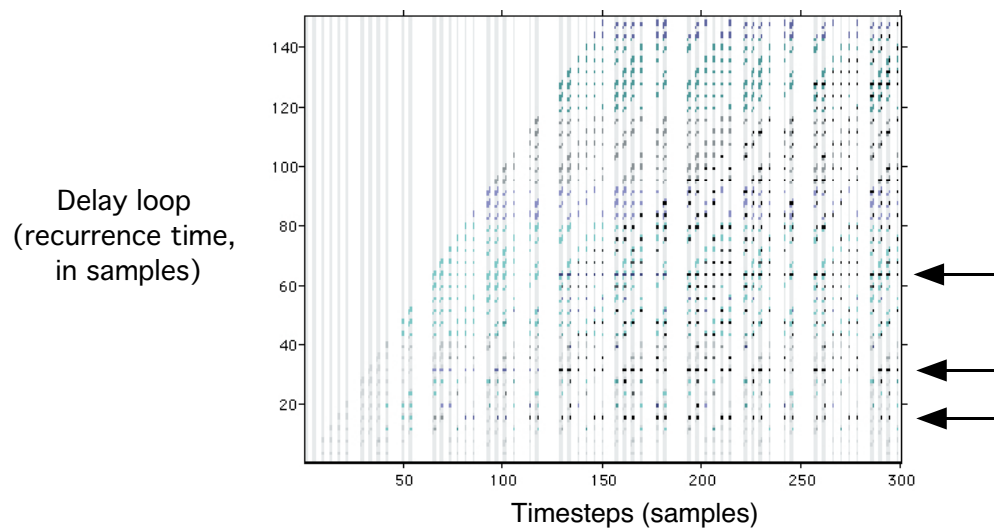


Figure 10. Behavior of recurrent timing nets. A. Behavior of a simple recurrent timing net for periodic pulse train patterns. The delay loop whose recurrence time equals the period of the pattern builds up that pattern. B. Response of a recurrent timing net to the beat pattern of *La Marseillaise*. Arrows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

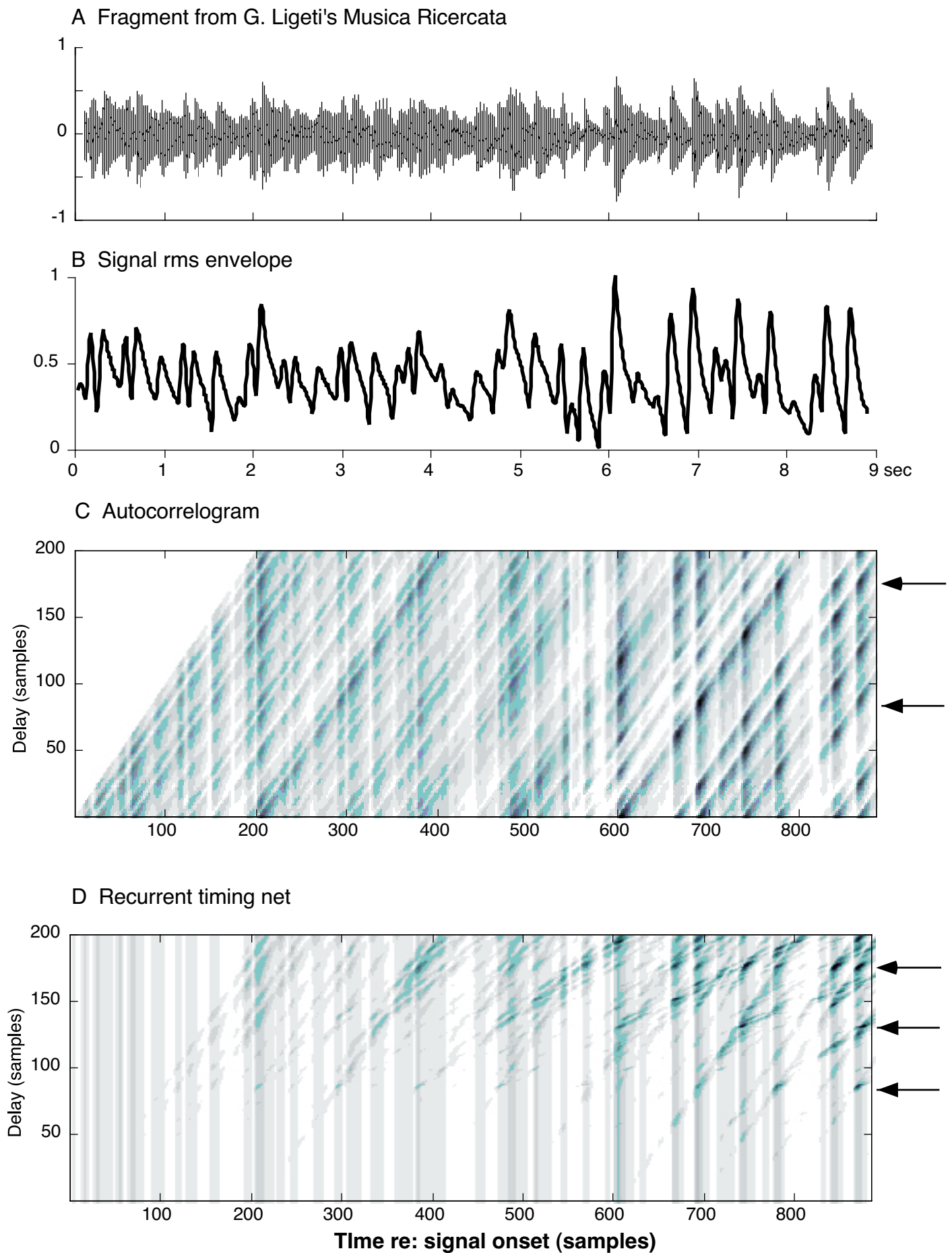
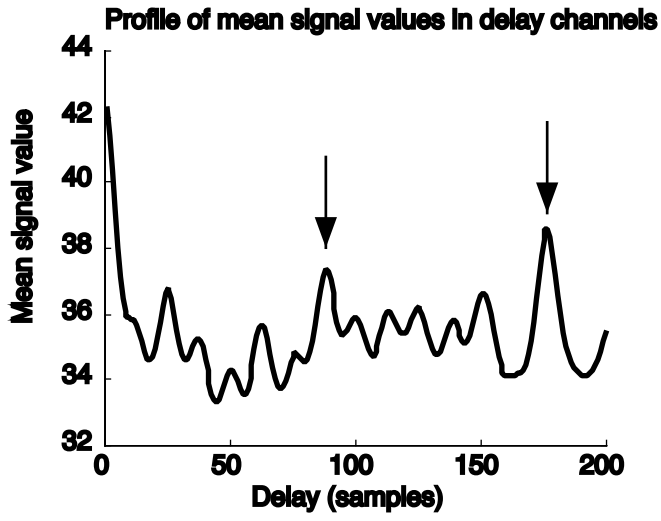
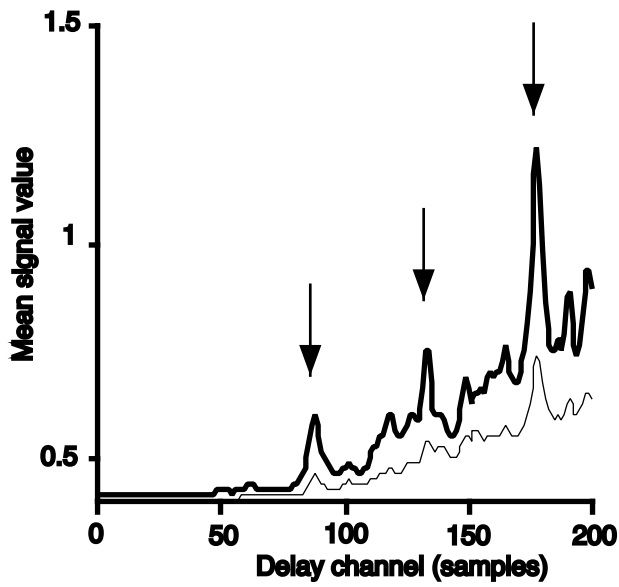


Figure 11. Analysis of complex rhythmic pattern in music. A. Waveform fragment from Ligeti, *Musica Ricercata*. B. Rms envelope of the waveform. C. Autocorrelogram (running autocorrelation) of the envelope. D. Response of the recurrent timing net. Arrows indicate delay channels that built up prominent patterns.

A Autocorrelogram



B Recurrent timing net



C Ligeti envelope (end)

