PERGAMON

Neural Networks 14 (2001) 737–753

Neural Networks

www.elsevier.com/locate/neunet

2001 Special issue

Neural timing nets

P.A. Cariani*

Eaton Peabody Laboratory of Auditory Physiology, Massachusetts Eye & Ear Infirmary, 243 Charles Street, Boston, MA 02114, USA

Received 10 October 2000; revised 22 March 2001; accepted 22 March 2001

Abstract

Formulations of artificial neural networks are directly related to assumptions about neural coding in the brain. Traditional connectionist networks assume channel-based rate coding, while time-delay networks convert temporally-coded inputs into rate-coded outputs. Neural timing nets that operate on time structured input spike trains to produce meaningful time-structured outputs are proposed. Basic computational properties of simple feedforward and recurrent timing nets are outlined and applied to auditory computations. Feed-forward timing nets consist of arrays of coincidence detectors connected via tapped delay lines. These temporal sieves extract common spike patterns in their inputs that can subserve extraction of common fundamental frequencies (periodicity pitch) and common spectrum (timbre). Feedforward timing nets can also be used to separate time-shifted patterns, fusing patterns with similar internal temporal structure and spatially segregating different ones. Simple recurrent timing nets consisting of arrays of delay loops amplify and separate recurring time patterns. Single- and multichannel recurrent timing nets are presented that demonstrate the separation of concurrent, double vowels. Timing nets constitute a new and general neural network strategy for performing temporal computations on neural spike trains: extraction of common periodicities, detection of recurring temporal patterns, and formation and separation of invariant spike patterns that subserve auditory objects.

Keywords: Neural timing networks; Time-delay neural networks; Temporal coding; Spiking neurons; Scene analysis; Temporal correlation; Auditory neurocomputation

1. Introduction

Traditionally, neural coding assumptions from neuroscience have informed the development of artificial neural networks. By far the predominant assumption has been that informational distinctions are encoded in profiles of average discharge rate across neurons, i.e. which neurons fire how frequently. Thus, which 'places' in cochleotopic, retinotopic, and somatotopic maps are activated have been thought to provide the basic information needed for form perception. In these neural networks the pulsatile, sequential character of spiking neurons is replaced by a continuously varying scalar quantity that reflects spike rate.

There have always been alternative, temporal theories of neural coding, however, in which information about the stimulus is conveyed via time patterns that the stimulus impresses on sensory neurons (Boring, 1942; Kiang, Watanabe, Thomas & Clark, 1965; Mountcastle, 1967; Troland, 1929; Wever, 1949). The pulse trains produced by spiking neurons are much more efficient transmitters of information encoded in relative timings of events rather than numbers of events (MacKay & McCulloch, 1952). It is in the functional context of processing temporally coded information, therefore, that neural architectures composed of 'spiking neurons' really come into their own.

Two broad classes of temporal codes stand out. Differences in temporal structure can arise through different times-of-arrival of spikes (latency- and synchrony-based codes) or through differences in temporal patterning of spikes (interspike interval and interval pattern codes). Different response latencies and patternings can be produced either by extrinsic, stimulus-locked responses of sensory receptors or through characteristic intrinsic temporal response patterns (e.g. different impulse responses). Thus there is a large space of possible neural pulse codes that can be based on which channels (labelled lines) are activated how much (rate-place codes), on the relative times-of-arrivals (latency codes), on spike patterning (temporal pattern codes), and even on joint response properties of particular subsets of neural elements (see Cariani, 1995, 1997b, 2001b; Mountcastle, 1967; Perkell & Bullock, 1968; Rieke, Warland, de Ruyter van Steveninck & Bialek, 1997; Sejnowski, 1999; Uttal, 1973; Wasserman, 1992).

In almost every sensory system, there exists temporal structure in neural response that is potentially capable of

* Tel.: +1-617-573-4243; fax: +1-617-726-5419.
E-mail address: peter@epl.meei.harvard.edu (P.A. Cariani).

0893-6080/01/$ - see front matter © 2001 Elsevier Science Ltd. All rights reserved.
PII: S0893-6080(01)00056-9

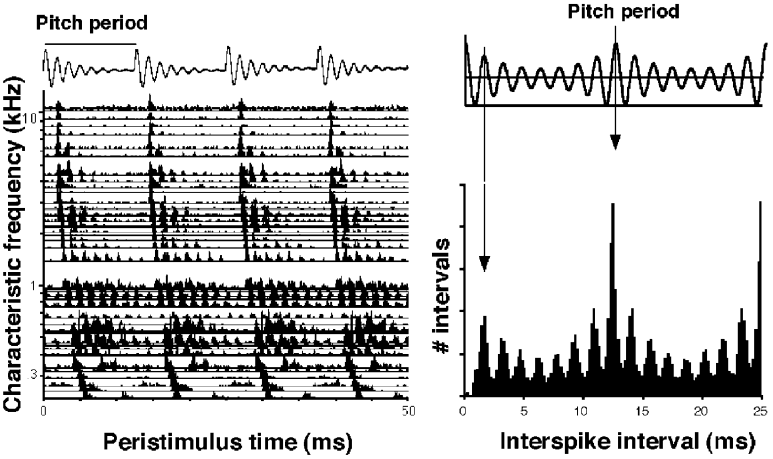

Fig. 1. Temporal coding of pitch and timbre in the auditory nerve. Top: Stimulus waveform, Single formant vowel, $F0 = 80$ Hz, $F1 = 640$ Hz, 60 dB SPL, 100 presentations/fiber. Left, peristimulus time histograms of the responses of 52 auditory nerve fibers of Dial-anestheized cats, arranged by fiber characteristic frequency. Top right, stimulus autocorrelation function. Bottom right, global, ensemble-wide distribution of all-order interspike intervals. The most frequent interval in the distribution is 12.5 ms, which corresponds to the stimulus fundamental period ($1/F0 = 1/80$ Hz) and the period of the low pitch that is heard. Other intervals correspond to periods related to formant-region partials that determine vowel quality (timbre) (see Cariani & Delgutte, 1996; Cariani, 1999).

supporting sensory quality distinctions (Cariani, 1995, 1997b; Mountcastle, 1967; Perkell & Bullock, 1968). Time structure related to stimulus quality exists in some rather unexpected places, such as in the chemical senses (Di Lorenzo & Hecht, 1993; Kauer, 1974; Laurent, 1999) and color vision (Kozak & Reitboeck, 1974; Young, 1977).

In many systems, phase-locked responses permit different response timings at different body locations to subserve localization functions (Carr, 1993; von Bekesy, 1967). At the behavioral level rather fine time-of-arrival disparities can be distinguished: fish electroception (<1 μs) microsecond; bat echolocation, (<1 μs to several μs), human interaural time differences (10–40 μs) (Colburn, 1996), and insect interaural time differences (1 ms) (Michelson, 1992). George von Bekesy reported human stimulus localizations based on as fine as 1 ms disparities in somatoception, olfaction, and gustation (von Bekesy, 1967). Motion detection in insect vision (Reichardt, 1961) and the limits of vernier acuity (Carney, Silverstein & Klein, 1995) may depend on comparisons of relative times of arrival in visual channels with different retinotopic locations. Spike precisions in visual systems on the order of a millisecond or less that could support fine spatiotemporal distinctions have been observed (Bialek, Rieke, van Stevenink & de Ruyter, 1991; Reinagel & Reid, 2000).

All of these perceptual computations are explicable in terms of temporal cross-correlation frameworks. The Jeffress model of binaural localization computed temporal cross correlations from phase-locked inputs (Jeffress, 1948). This model was one of the very first neural networks to successfully account for specific aspects of perception, and it inspired subsequent models in other sensory modalities.

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 is used to subserve the representation of a number of auditory qualities: pitch, timbre, rhythm, sound location. The case of pitch is illustrative. Robust and pervasive correspondences between patterns of human pitch judgment and the global all-order interval statistics of populations of auditory nerve fibers have been found in models, simulations and neurophysiological studies (Cariani, 1999a; Lyon & Shamma, 1995; Meddis & Hewitt, 1991a; Slaney & Lyon, 1993). Features of these population–interval distributions (Fig. 1) closely parallel human pitch judgements (Cariani & Delgutte, 1996): the most frequent all-order interval corresponds to the pitch that is heard, and the fraction of this interval amongst all others corresponds to its strength (salience). 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. Timbres of stationary sounds such as vowels correspond to distributions of short (<5 ms) interspike intervals.

As a direct consequence of phase-locking, positions of major and minor peaks in observed population–interval distributions closely mirror those of their respective stimulus autocorrelation functions. For complex stimuli with unresolved harmonics (e.g. >2 kHz), population–interval distributions reflect waveform envelopes. These distributions thus provide general-purpose autocorrelation-like representations for stimulus periodicities up to the limits of robust phase locking (~5 kHz). Rather than the temporal cross-correlations that subserve localization, temporal autocorrelations appear to subserve the computation of auditory forms. The first neurocomputational model to compute temporal autocorrelations to explain the pitches produced by complex tones was J.C.R. Licklider's time-delay 'duplex' network (Licklider, 1951, 1959).

2. Neural codes and neural networks

The near ubiquity of spike timing information in sensory processing begs the question of what kinds of neural architectures are generally needed to make use of it. Each kind of neural code requires a correspondingly different kind of neural network for its analysis. If one considers the basic division between both channel-based, rate–place codes and temporal codes, four basic transformations are possible (place–place, place–time, time–place, and time–time). If networks require differential weightings of connections to distinguish different activation patterns and time delays to distinguish differences of timing, then three broad classes of networks are created: connectionist networks, time-delay networks, and timing nets (Table 1).

Connectionist networks operate on across-element discharge rate profiles in their inputs to produce meaningful rate–coded output patterns (place–place transformations). By far the majority of neural network research has focused on feedforward, recurrent, and adaptive connectionist networks.

Time-delay networks (TDNNs) traditionally transform temporally-coded inputs into rate-coded outputs by incorporating inter-element time delays as well as connection weights. The Jeffress model of binaural localization and the Licklider model of pitch perception were auditory time-delay networks that transformed temporally coded inputs into spatial activation profiles in order to compute temporal autocorrelations (for pitch) and cross-correlations (for binaural localization). These models used coincidence counters that combined coincidence detection with a subsequent integration (counting) process. Such architectures transform the fine time patterns in their inputs to smoothed, running averages of numbers of spike coincidences. Other implementations of time-delay networks use arrays of oscillators rather than delay lines and coincidence counters to discriminate different periodicities (Wang, 1995). In both kinds of implementations there is an explicit measurement of periodicity that is associated with each particular element, such that the output of the network is an across-element profile of activated elements. Time-delay networks can also be used to effect place–time transformations, producing characteristic output time patterns when particular spatial patterns of activation are presented (as in central pattern generators and in oscillator-networks that synchronize on the basis of spatial input patterns).

Recently we have proposed another class of neural nets, called *timing nets*, that operate on temporally-coded inputs to produce meaningful temporally-coded outputs (Cariani, 2001a). This paper discusses some of the basic computational properties of simple feedforward and recurrent timing nets. Many of the basic properties presented here were also outlined in that paper.

Much of our motivation for exploring the properties of such networks has been driven by the quest for an explanation of how the auditory system uses interspike interval information for the computation of pitch. One needs to explain how the auditory system is capable of reliably making very fine pitch distinctions ($<1\%$ in frequency) over very large dynamic ranges (>80 dB) using neural elements that are, in comparison with the percept, relatively coarsely tuned. This is the persistent 'hyperacuity problem' that currently exists for many sensory qualities (Rieke et al., 1997). In the auditory system, frequency hyperacuity is especially a problem at high sound pressure levels, where rate-based frequency tunings broaden dramatically, but perception remains precise. As a consequence of this broadening, there are fundamental difficulties in accounting for the precision and robustness of frequency discriminations in terms of average discharge rates. However, interval information, like frequency discrimination, remains exceptionally precise over the entire dynamic range. Frequency discrimination covaries with the amount of interval based information, such that interval-based representations of frequency account well for the decline in frequency discrimination as frequency increases and phase-locking weakens (Goldstein & Srulovicz, 1977). As was noted above, the interval patterns also explain an exceptionally wide range of complex, subtle, and unexpected patterns of pitch judgements.

Currently most auditory physiologists believe that a time–place transformation is effected in the auditory pathway by neural elements that are tuned to particular periodicities (Langner, 1992). However, tunings of these elements in the auditory pathway are coarse in comparison the pitch distinctions they are thought to subserve. Further, these tunings broaden at higher sound pressure levels (Krishna & Semple, 2000; Rees & Møller, 1987). There are other problems with such accounts that have to do with differences between autocorrelational representations and those based on modulation spectrum. Perception of pitches produced by perceptually-resolved, lower-frequency harmonics, for example, follows an autocorrelation-like analysis rather than a modulation-based analysis of waveform envelopes (de Boer, 1976). If a time–place transformation were effected by the auditory system, then the

Table 1
General types of neural networks

Type of network	Inputs	Outputs
Connectionist network	Spatial excitation patterns	Spatial excitation patterns
Time delay network	Temporal spike patterns	Spatial excitation patterns
Timing net	Temporal spike patterns	Temporal spike patterns

elements ideally should carry out a temporal autocorrelation analysis, which would require that they have characteristics of comb-filters. Unfortunately, thus far no such elements with these characteristics have been observed, so that arguably, there exist no strong neural candidates for the pitch detectors that a time-to-place account would require. The absence of precise and robust pitch detectors notwithstanding, interspike interval distributions at early stages of auditory processing do retain the requisite properties for neural substrates of pitch (Cariani, 1999a). As a result, alternative neurocomputational strategies that retain the information in the time domain have been explored.

3. Feedforward timing networks

Alongside traditional connectionist networks and time-delay networks, neural timing networks can be envisioned that operate on time structure in their inputs to produce interpretable temporal patterns in their outputs (Cariani, 2001c). These networks consist of coincidence detectors and delay lines which analyze temporally-coded inputs. Their closest precursors are simple functional models of neural computation for which fine time structure is of primary importance (Abeles, 1990; Braitenberg, 1961, 1967; Jeffress, 1948; Longuet-Higgins, 1989; MacKay, 1962; Reitboeck, 1989; Thatcher & John, 1977). Some aspects of timing nets were inspired by the functional anatomy of cortical structures (Braitenberg, 1961; Reitboeck, 1989; Thatcher & John, 1977), while others were inspired by signal-processing operations that they elegantly implement (Braitenberg, 1961; Cherry, 1961; Longuet-Higgins, 1987, 1989). There are also a number of time-domain auditory processing models that operate on phase-locked spike timing information to produce spatial patterns of activation that serve as 'central spectrum' representations. Some of these operate on local synchrony, either using coincidence-based (Young & Sachs, 1979) or cancellation-like operations (Colburn, 1996; Colburn & Durlach, 1978; Culling, Summerfield & Marshall, 1998; de Cheveigné, 1998; Seneff, 1985, 1988). Other time-domain analyses use global synchronies between non-neighboring frequency channels as a means of implementing harmonic templates for spatial-pattern representation of the pitches of harmonic complex tones (Shamma & Sutton, 2000). Analogously, early time-domain models operated on interspike interval information within each frequency channel (Licklider, 1951, 1959; Lyon & Shamma, 1996), while later interval-based models formed global temporal representations that retain no 'place' information (Ghitza, 1988; Lyon, 1984; Meddis & Hewitt, 1991a,b; Meddis & O'Mard, 1997; Moore, 1997; van Noorden, 1982). Temporal correlation models for binaurally-created pitches (Akeroyd & Summerfield, 1999; Cariani, 1996, 2001a) are the closest existing implementations to the feedforward nets presented here, in that the global temporal structure of coincidences in output of binaural cross-correlation arrays carries the information that determines the pitch. Hypothetical neural timing architectures guide and are also guided by correlation-based analyses of neural function (Abeles, 1990; Eggermont, 1990, 1993; Johannesma, Aertsen, van den Boogaard, Eggermont & Epping, 1986). While the motivation in this paper is primarily pragmatic, to see what useful signal processing functions timing nets afford, our ultimate aim is scientific, to widen the range of neurocomputational hypotheses that are available to us we attempt to reverse-engineer the brain.

Consider an array of coincidence detectors that have inputs from two sets of tapped delay lines arranged in anti-parallel orientation [Fig. 2(a)]. Two spike trains are fed in from either end of the array and propagate through their respective delay lines. Spikes in the two trains cross at different points in the array; when there is simultaneous arrival of spikes in both channels, the coincidence detector at the crossing point produces an output spike (depicted in the figure by spikes on the output lines below the detectors). Many relative delays are realized by the slow conduction times across the array such that each position along the tapped delay line corresponds to a particular relative delay between the input signals (D_{ij}). Since each coincidence detector with a relative delay D_{ij} requires nearly simultaneous arrival of a spike in both lines in order to fire, each spike in the output of the coincidence array represents the joint occurrence of spike arrivals in the two inputs [Fig. 2(b)]. If spike trains are represented by binary-valued (0,1) time series S_i and S_j, (spike occurrence at time $t = 1$, 0 otherwise), then the output of a particular coincidence detector C_k is $S_i(t)$*$S_j(t-D_k)$.

3.1. Basic computations

Several basic computations can be carried out. First, the cross-correlation function (CCF) of the two inputs can be computed by counting the number of spikes in each output channel (vertical gray bar) as a function of relative delay D_k. $\mathrm{CCF}(D_k) = \Sigma S_i(t) * S_j(t - D_k)$, summed for each delay channel over all times t ($*$ denoting multiplication). Their convolution can be computed by summing across relative delay channels for each time step (horizontal gray bar). $\mathrm{Conv}(t) = \Sigma S_i(t) * S_j(t - D_k)$, summed over all D_k for each time step. The operation is similar to the common flip/shift/multiply method of computing convolutions. In terms of spike train analysis, this would be a population peristimulus time histogram. The distribution of all-order interspike intervals in the output of channel C_k is the same as the autocorrelation function (AC) of the output spike train, i.e. if τ is an interspike interval duration or time lag, then $\mathrm{AC}_k(\tau) = \Sigma S_i(t) * S_j(t - D_k)] * \Sigma S_i(-\tau) * S_j(t - \tau - D_k)]$ over all times t. The summary–autocorrelation or population–interval distribution of the outputs is the sum of the autocorrelations of each of the output channels, $\mathrm{SAC}(\tau) = \Sigma \mathrm{AC}_k(\tau)$. The population–interval distribution

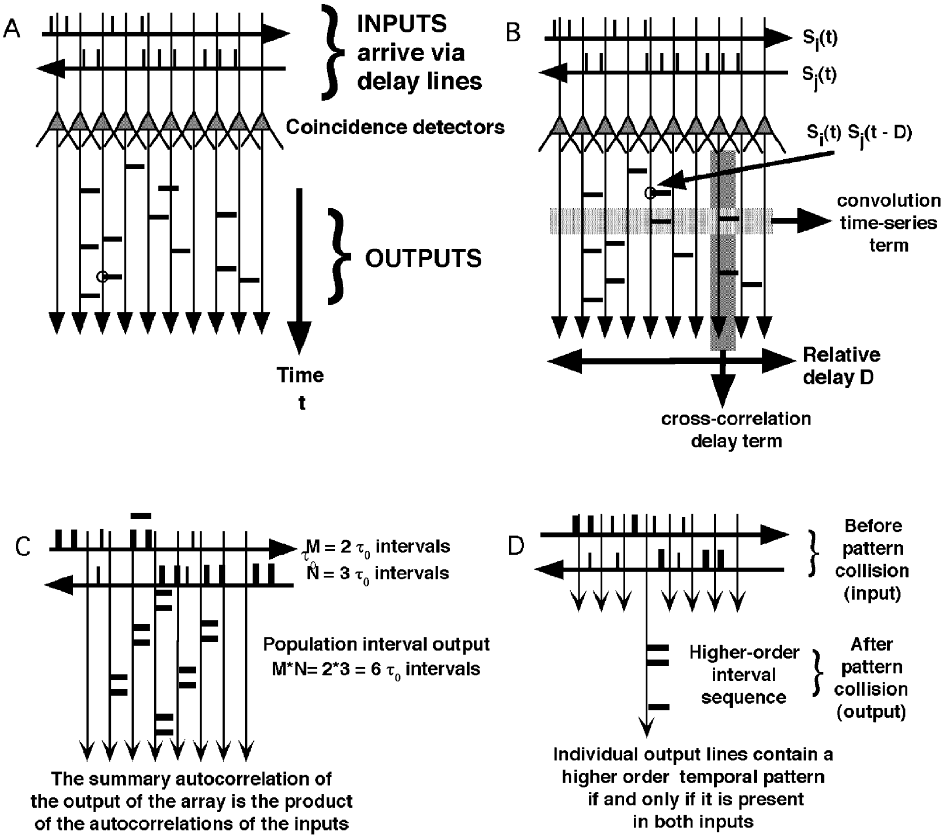
A
INPUTS arrive via delay lines
Coincidence detectors
OUTPUTS
Time t
B
$S_i(t)$
$S_j(t)$
$S_i(t)\ S_j(t - D)$
convolution time-series term
Relative delay D
cross-correlation delay term
C
M = 2 τ_0 intervals
N = 3 τ_0 intervals
Population interval output
M*N= 2*3 = 6 τ_0 intervals
The summary autocorrelation of the output of the array is the product of the autocorrelations of the inputs
D
Before pattern collision (input)
Higher-order interval sequence
After pattern collision (output)
Individual output lines contain a higher order temporal pattern if and only if it is present in both inputs

Fig. 2. A simple feedforward timing net. (a) General schematic of a coincidence array traversed by tapped delay lines. Each coincidence element (pyramidal cell icons) receives two inputs. The spatial position of a coincidence element in the array determines the relative time of arrival of the two inputs (delay D). When pulses in the two lines arrive simultaneously, an output pulse is emitted. (b) Convolution and cross-correlation functions. 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. The conduction time across the array enforces a temporal contiguity window for signal comparisons. (c) The population–interval (summary autocorrelation) of the entire output ensemble computes the product of the autocorrelations of the two input channels. (d) Intervals and higher-order interval patterns will appear in the individual output lines only if such patterns are present in both input lines. Such higher-order patterns will appear in individual output lines even in the face of embedded spikes that are not part of the pattern.

is the distribution of all-order interspike intervals that are produced by the coincidence array. This is the global temporal neural representation that was noted above to account for many aspects of pitch and timbre, and this is the output representation that will be utilized.

The finite spatial extent of the array enforces a temporal contiguity constraint on the operations. If the conduction time across the array is 20 ms, then only those portions of the two input spike trains that are produced within the same 20 ms window will cross in the array. All interspike intervals whose constituent spikes do arrive within this temporal contiguity window will cross their counterparts in the other set, such that if one input has M intervals of duration τ, and the other has N such intervals, M*N τ-length intervals will appear in the outputs [Fig. 2(c)]. The coincidence array therefore performs a multiplication of the autocorrelations of its inputs, taking into account the contiguity window.

A further consequence is that each interspike interval or higher-order spike arrival pattern, such as a triplet, appearing in a given output channel must also be present in each of the two inputs [Fig. 2(d)]. The array thus functions as a temporal sieve, passing those temporal patterns that are common to both sets of inputs. Such complex patterns will appear in the output of individual channels even if they are embedded in other spikes. Thus if one wants to compute whether a given spike train contains a pattern, one generates the pattern of interest and feeds it into the coincidence net. If the pattern is present in the other input, then it will reappear in the output. This is potentially relevant to temporal multiplexing by means of different interleaved time patterns (e.g. Cariani, 1997a; Emmers, 1981). Here the presence of a particular subpattern can be detected amidst many others. This affords modes of multiplexing that are akin to 'code-division multiplexing' in which different temporal patterns asynchronously convey different signals (Cariani, 1997a; Chung, Raymond & Lettvin, 1970). This is somewhat different from the 'time-division' multiplexing schemes that are more often proposed (Singer, 1995). The output of such a pattern-detection process can be iterated and/or fed back on itself such that more copies of the input pattern are produced. This becomes a means by which particular patterns can be amplified by such systems.

Structurally, this architecture is reminiscent of both the Jeffress binaural localization model (Jeffress, 1948) and the Braitenberg cerebellar timing model (Braitenberg, 1961). The present architecture differs from these and most time-domain auditory models in its functioning. Here, in contrast with those models, no subsequent ‘counting’ or rate integration stage is included, since the output of this network is the time structure it produces in its inputs rather than an across-element activation pattern. The nature of the output signals involved is thus very different. Timing net computations consequently bear greater resemblance to analog signal processing operations (Mead, 1989) that produce time-series analog outputs than to digital signal processing algorithms that produce explicit representations, be they profiles of numerical parameter values, feature-detections, or element activations.

3.2. Extraction of common periodicities

Coincidence arrays extract those periodicities 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). On longer time scales, rhythms can be compared to detect common underlying meters and subpatterns.

As a simple example, two amplitude modulated tones having the same fundamental frequency ($F0 = Fm = 125$ Hz), but different carrier frequencies ($Fc = 500$ vs 1250 Hz) were synthesized (Fig. 3). Simulations were carried out in MATLAB. Perceptually, these two signals produce the same low pitch at their common ‘missing’ fundamental, $F0 = 125$ Hz, despite the lack of any common partials in their power spectra (middle plot). The signals, constructed with a 10 kHz sampling rate, were half-wave rectified. Waveform maxima were replaced by rectangular pulses 300 μs wide. Sample spike trains are shown above their respective signals. Crudely, these resemble the phase-locked responses of auditory nerve fibers, albeit at some higher sustained firing rates than would be seen physiologically in individual units. In real neural systems, synchronized discharges across several neurons would be required to support representation of higher periodicities through a ‘volley principle’ (Wever, 1949). All-order interspike interval histograms of the 100 ms spike trains are shown on the left. The spike trains share a common periodicity at the fundamental period ($1/F0 = 8$ ms).

The pulse trains were passed through the coincidence network. Coincidences produced by the network is shown in the bottom left panel. The population interval distribution was computed by summing together the all-order interspike interval distributions of each of the output channels. Intervals corresponding to the common fundamental period of 8 ms dominate the output of the network. The coincidence array thus passes into individual output channels only those temporal patterns that are common to the two inputs. Effectively, the network extracts the common fundamental periodicity without ever making any sort of explicit estimation of the fundamentals of the two input signals.

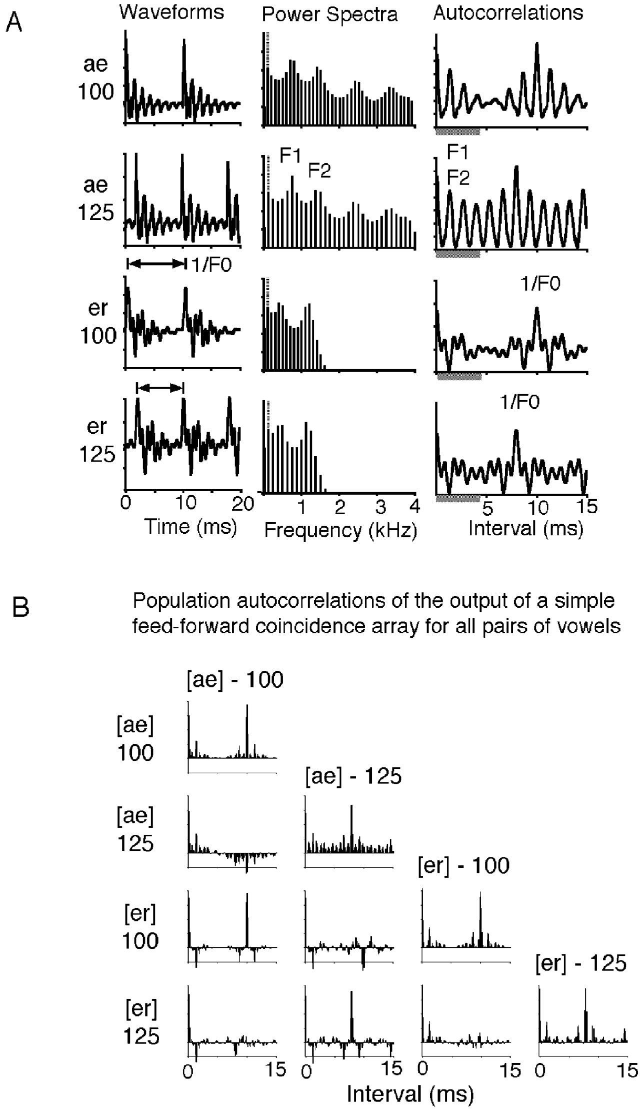

Fig. 3. (a) Waveforms, power spectra, and autocorrelation functions for four vowels. The vowel set consists of combinations of two different fundamental frequencies ($F0 = 100$, 125 Hz) and two formant structures. Horizontal arrows above waveforms and vertical lines in autocorrelations indicate fundamental periods ($1/F0 = 8$, 10 ms), which correspond to voice pitch periods. Shaded bars indicate periodicities associated with formant structure that give rise to differences in vowel quality (timbre). (b) Population autocorrelations of the output of the coincidence array for all vowel pairs.

Coincidence nets can also extract common periodicities that are associated with different timbres or vowel qualities (Cariani, 2001a). This is useful for the extraction of common pitches irrespective of differences in timbre (e.g. the same musical instrument playing different notes, or two different people speaking the same vowel). Four synthetic vowels consisting of combinations of two fundamental frequencies and two spectral envelopes (formant combinations) were constructed [Fig. 4(a)]. These particular synthetic vowels

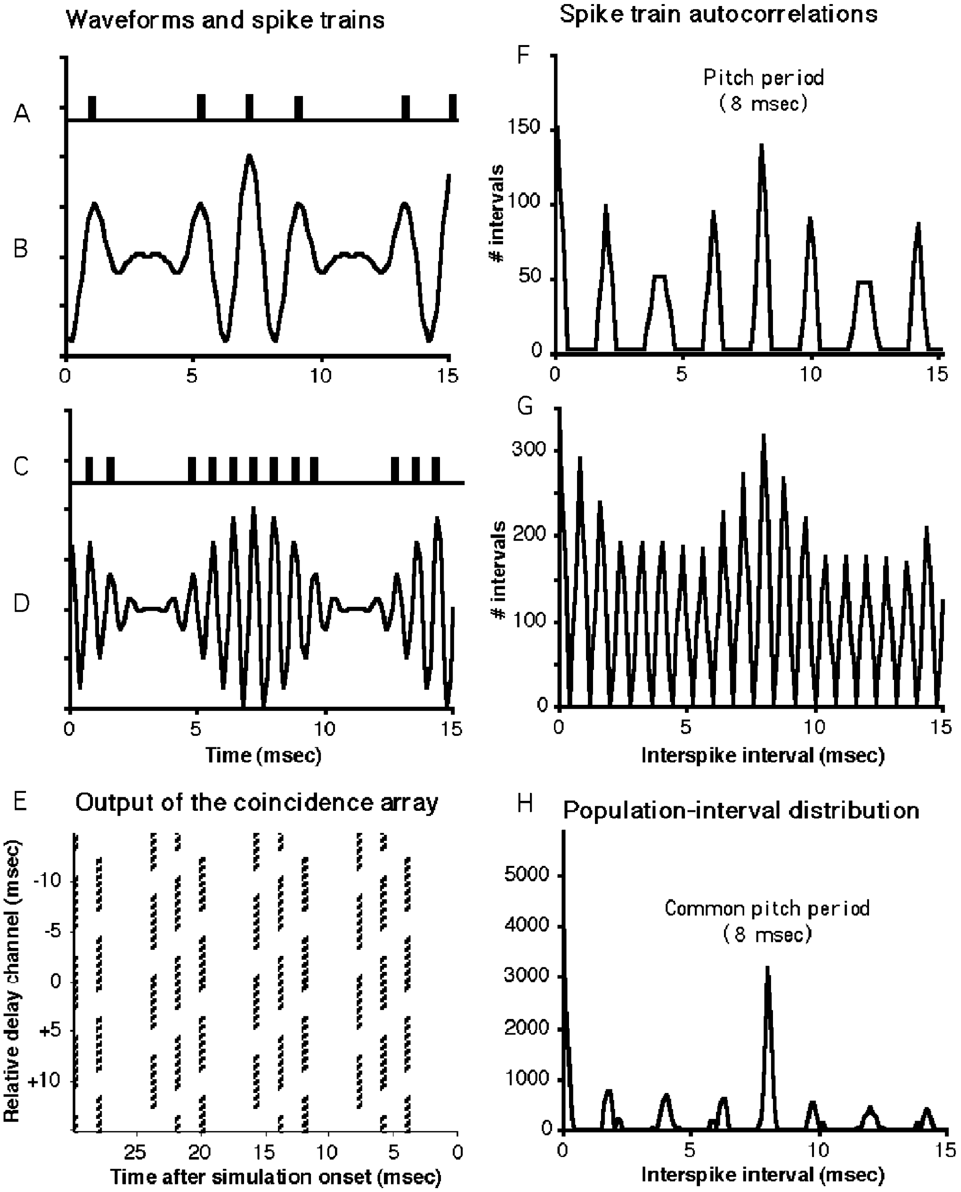

Fig. 4. Extraction of a common fundamental frequency by a coincidence array. (a) and (c) Pulse trains derived from maxima of two waveforms. Pulses are 300 μs wide. (b) and (d) Waveforms of two amplitude-modulated (AM) tones with different carriers ($Fc = 500$, 1250 Hz) but same modulation frequency ($Fm = 125$ Hz). The AM tones have no harmonics in common, but have a common fundamental frequency that produces the same low pitch at their 'missing fundamental' ($F0 = Fm = 125$Hz). (e) Output of a coincidence array in response to presentation of the two pulse trains. The effective coincidence window was 600 μs. The output of the coincidence array is shown as a function of the relative delay channel (ordinate) and time (abcissa). (f) and (g) The all-order interspike interval distributions of the input pulse trains. (h) The population–interval distribution of the output of the coincidence array.

most closely correspond to the vowels [ae] (a as in hat) and [er] (er as in herd). Waveforms, power spectra, and autocorrelations of the vowels are shown. The signal-property correlates of the voice pitches that are heard are (1) the period of temporal pattern in their waveform, (2) the harmonic spacings in their spectra, and (3) the positions of major peaks in their autocorrelation functions (vertical lines). The correlates of their vowel quality or timbre that distinguish them as different phonetic-types are (1) the internal structure of the repeating waveform pattern, (2) the shape of their spectral envelopes, and (3) patterns of short intervals (up to half the fundamental period) in their autocorrelations (bars under the plots). Phonetic identities of different vowels can thus be distinguished on the basis of waveforms, power spectra, or autocorrelations. Population–interval distributions at the level of the auditory nerve provide effective autocorrelation-like neural representations of vowel identity (Cariani, 1995, 1999a; Cariani & Delgutte, 1993, 1994; Cariani, Delgutte & Tramo, 1997; Lyon & Shamma, 1996; Palmer, 1992) whose features closely follow phonetic boundaries (Hirahara, Cariani & Delgutte, 1996a,b). The positions of minor peaks in population–interval distributions estimated from neural

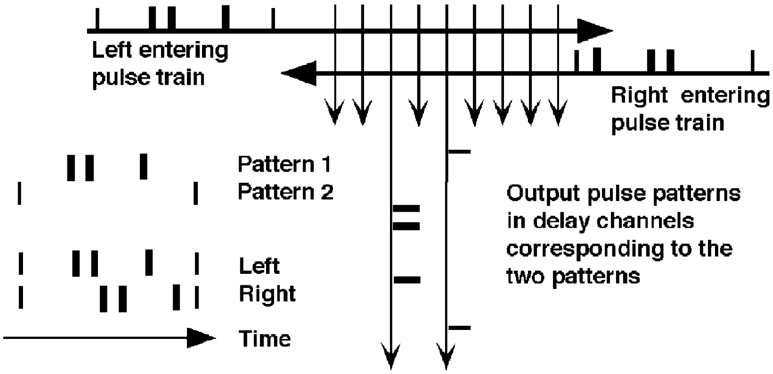
Left entering pulse train
Right entering pulse train
Pattern 1
Pattern 2
Left
Right
Time
Output pulse patterns in delay channels corresponding to the two patterns

Fig. 5. Fusion and separation of patterns by relative arrival time disparities in a feedforward timing net. Two temporal patterns of pulses are shown that have different inter-channel arrival times, such that the temporal disparity between the two patterns is different in each channel. As a consequence of this difference, the two patterns cross their counterparts at different relative delays and are propagated by two different delay channels.

data mirror those in the stimulus autocorrelation function. For example, the patterns of minor interval peaks in the population–interval distribution in Fig. 1 are associated with the formant frequency of the single-formant vowel stimulus.

Thus far, we have dealt with binary-valued pulsatile signals. Operations on binary pulse trains can be scaled up to handle positive-real-valued signals by assuming many of the same operations are carried out in parallel using spike probabilities amongst many neural elements, e.g. peristimulus time (PST) histograms. For example, if a stimulus generates 50 spikes at time t amongst 100 neural elements, then 50 spike trains out of 100 in a parallel array will have a spike at time t. If a second stimulus generates 20 spikes out of 100 neural elements in a second population, then 20 spikes will occur at time t for that population. If all 10,000 pairwise combinations of spike trains generated by the two populations are fed into coincidence detectors, then $20 * 50 = 1000$ coincidences will be produced. The probability of a coincidence in any one pair is 1000/10 000, or 0.1. This is equal to product of the respective probabilities for $S_i(t) = 1$ and $S_j(t) = 1$, i.e. the expected relative frequency of coincidences in the ensemble at time t is $0.5*0.2 = 0.1$.

All four synthetic vowels were passed pairwise through a coincidence net. The summary autocorrelation of the output of the net for each pair is shown in [Fig. 4(b)]. Those vowel pairs that had common fundamentals (the same voice pitches: ae-100, er-100 and ae-125, er-125), when passed through the network produced summary autocorrelations with major peaks at their common fundamental periods (their common pitch). Those vowel pairs that had common formant-structure (the same vowel but different voice pitch: ae-100, ae-125 and er-100, er-125), produced patterns of short intervals that corresponded to their respective formant structures. Those vowel pairs that had neither common fundamental frequency nor common formant structure (different voice pitch and different timbres: ae-100, er-125 and er-100, ae-125), produced only small peaks associated with overlapping subharmonics. Thus the same mechanism handles both pitch and timbre, extracting similarities of pitch irrespective of timbre and similarities of timbre irrespective of pitch.

Comparable results were also obtained using multichannel auditory nerve simulations that incorporated bandpass filtering, half-wave rectification, low pass filtering, and rate compression (Cariani, 2000). The results are similar because phase-locking produces spike trains whose time structure resembles the waveform and whose autocorrelations (all-order interval histograms) resemble the positive portions of the stimulus autocorrelation functions. In simulations and in observed neural population responses, nonlinear processes, such as floor and saturation of discharge rates, alter relative heights of interval peaks without modifying the periodicities that the peaks represent (Cariani, 1999a; Cariani et al., 1997). The functional effects of nonlinearities thus depend on the nature of the neural codes involved. Interspike interval codes for representation of periodicity are highly resilient to the effects of such nonlinearities.

3.3. Separation of auditory objects by means of time-of-arrival disparities

Coincidence networks also can play a role in building up fused images and separating them on the basis of time-of-arrival disparities (Fig. 5). This was originally suggested by Colin Cherry's model for binaural fusion (Cherry, 1961), which combined autocorrelation and cross-correlation operations. When two low-frequency tones are presented with different interaural time disparities (ITD), they are heard at different azimuthal locations. Consider, however, the behavior of a coincidence array when two different pulse sequences (Patterns 1 and 2) with two different temporal disparities are presented to the network. Because of the temporal disparity between patterns, each pattern meets its counterpart at a different relative delay. Consequently, the two temporal patterns are segregated into two different output channels. This kind of mechanism works for both interaural time disparities and onset time disparities, albeit in different delay regimes. Both cues play a role in the perceptual fusion of sounds.

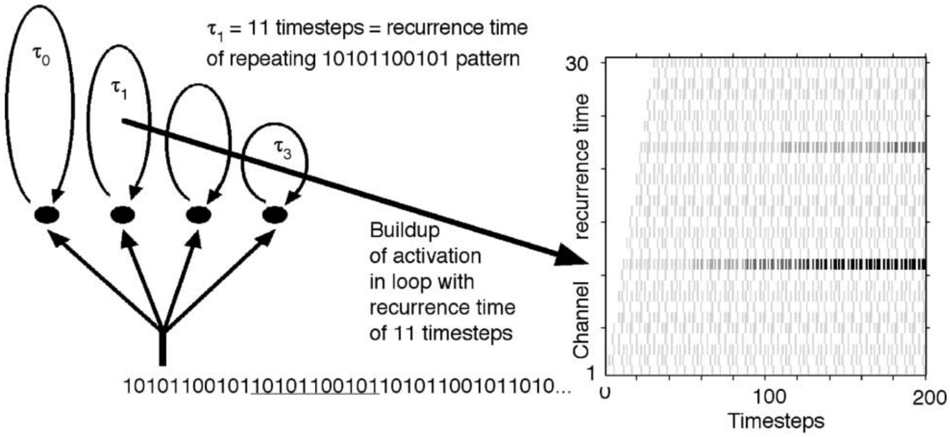

Fig. 6. Behavior of a simple recurrent timing net for periodic pulse train patterns. Left, recurrent timing net consisting of an array of coincidence detectors with associated loops having a range of recurrence times. Right, output of the coincidence array, arranged by the length of delay loop (ordinate) and time (abcissa). Periodic patterns invariably build up in the delay loop whose recurrence time equals the period of the pattern (from Cariani, 2001a).

In audition, interaural disparities are interpreted as azimuthal locations. In vision, depth cues are created by both binocular spatial disparities and temporal disparities (e.g. the Pulfrich effect). It is conceivable that such temporal processing could be applied to problems of stereopsis in binocular vision. Application to binocular fusion and depth perception would require a conversion from spatial to temporal pattern, i.e. a scanning process (Pabst, Reitboeck & Eckhorn, 1989; Reitboeck, Pabst & Eckhorn, 1988). Temporal correlations between retinal channels might be obtained from horizontal image motion coupled with phase-locking of retinal elements to edges. Provided with such temporal substrates, a simple coincidence net would fuse and segregate binocular images in a manner similar to the processing of binaural images.

3.4. General implications

An important general property of these feedforward timing networks is that their functioning depends neither on particular interconnections nor on which particular elements are activated. As long as there are rich sets of delays, for purposes of pattern extraction, these networks are indifferent as to which particular coincidence elements are activated (for purposes of localization, as in Fig. 5, coincidence arrays do need to be ordered, and on which coincidence channel the output patterns appear does matter). Populations of neurons connected by means of these coincidence nets therefore could potentially process information asynchronously, in mass-statistical fashion. Since they operate on interval statistics that do not depend on the particular transmission channels involved, provided there are many relative delays, such networks may obviate the need for precise point-to-point connectivities. This in turn may permit information to be broadcast en masse, without having to guarantee in advance a coherent constellation of path- and element-specific connection weights and conduction times.

4. Recurrent timing nets

The comparisons outlined above require the two sets of inputs to be simultaneously present in the network in order to beat them together. Consequently, for delayed matching tasks, timing information must be stored and retrieved. The simplest temporal storage strategy is to allow the signals themselves circulate in a reverberating conduction loop, as temporal memory traces. Incoming time patterns can then be compared with circulating ones using the kinds of correlational operations outlined above. Stimulus matching in such a system would entail maximizing the output of the whole coincidence array. In addition percepts build up over time, with previous patterns dynamically creating sets of perceptual expectations that can either be confirmed and built up or violated. Periodic signals, such as rhythms, thus build up their own temporal expectations. These recurrent timing networks were inspired in different ways by the stabilized auditory images of (Patterson, Allerhand & Giguere, 1995), the neural loop model of (Thatcher & John, 1977), the adaptive timing nets proposed by (MacKay, 1962), the adaptive resonance circuits of (Grossberg, 1988), and the psychology of temporal expectation (Jones, 1976; Miller & Barnet, 1993). With these ideas in mind, computational properties of simple recurrent timing nets were explored.

4.1. Buildup of periodic time patterns

The simple recurrent timing network in Fig. 6 cross-correlates incoming time patterns with previous, circulating ones in order to build up those temporal patterns that recur. As a first step, pulse trains with repeated, randomly selected pulse patterns (e.g. 100101011–100101011–100101011...) were passed through the network. The

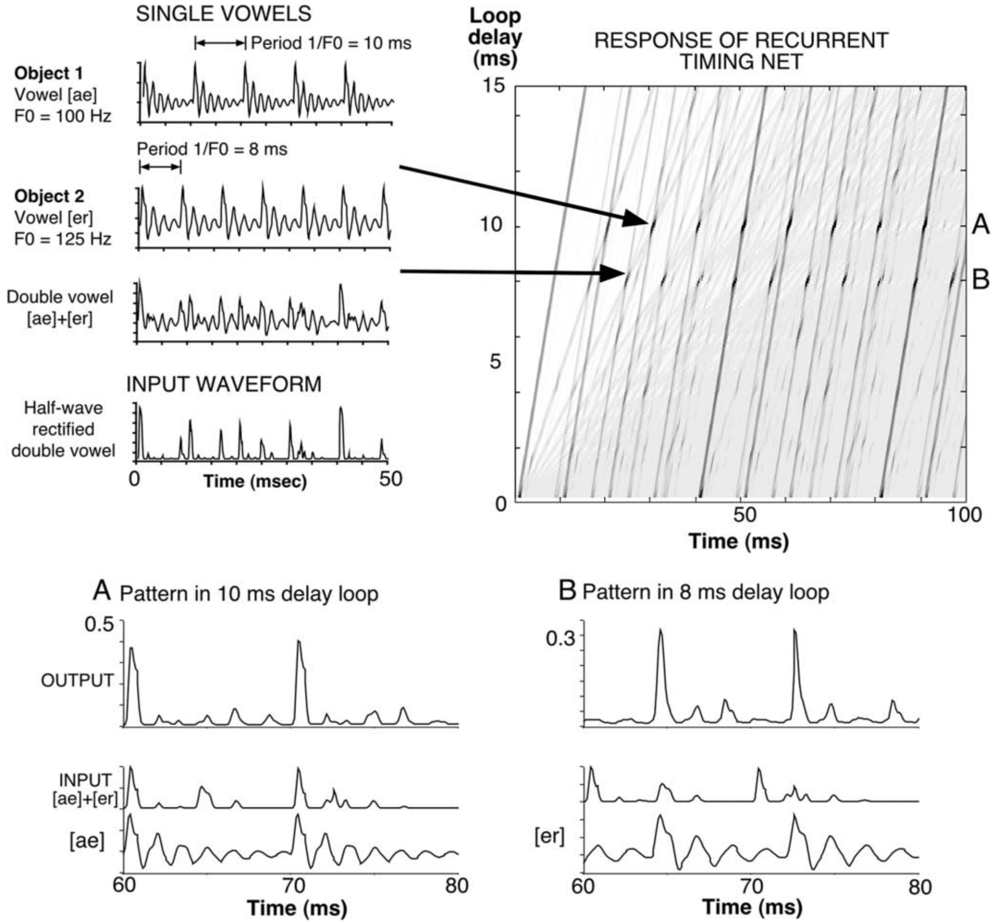

Fig. 7. Separation of auditory objects through temporal pattern coherence. Top waveform plots. Two synthetic vowels [ae] and [er] with different fundamental frequencies 100 and 125 Hz were summed together to form a double vowel. The composite waveform was half-wave rectified and presented to a recurrent timing net. Right top, the output of the recurrent coincidence array shows the buildup of the two patterns in the delay channels (A and B) whose recurrence times equal their respective periods. (a) Top, waveform of the signal circulating in the 10 ms delay loop at 60–80 ms after the stimulus onset. Middle, corresponding input waveform for the same time period. Bottom, waveform of the vowel [ae] that has a fundamental period of 10 ms for the same period. (b) As with (a), except that the vowel [er] is plotted.

same input signal, here a pulse train, is presented to each of the coincidence elements in the array. For each time step, the incoming unit-amplitude pulse train is multiplied by the variable-amplitude train arriving in the delay loop. In the absence of pulses arriving through the delay loop, the incoming unit-amplitude pattern is fed into the loop. Coincidences between pulses increase the amplitudes of pulses that propagated back through the loop by 5%.

The plot shows the signals produced by the coincidence elements as a function of their loop delay (ordinate) and time (abcissa), with signal strength being indicated as shades of gray (black is maximal, white is zero). Here the pattern-period is 11 timesteps, and the delay loop that builds up the strongest signal has a recurrence time of 11 timesteps. Irrespective of the pattern that is repeated, periodical pulse patterns invariably build up fastest in the delay loop whose recurrence time matches their repetition time. Thus, recurrent time patterns are repeatedly correlated with themselves to build up to detection thresholds. In effect, the cross-correlation loops dynamically create matched filters from repeating temporal patterns in the stimulus. In this manner, temporal-pattern invariances are enhanced relative to uncorrelated patterns. In essence, the network functions as a pattern-amplifier. Related kinds of correlation-based strategies were used in the 1950s to detect periodic signals in noise (Lange, 1967; Meyer-Eppler, 1953), in situations where the period of the target signal was known a priori. This network implements such periodicity-amplification and detection strategies in a more systematic and general way.

4.2. Formation and separation of auditory objects through temporal coherence

When two repeating temporal patterns each with its own repetition period are summed and presented to a recurrent timing net, the two patterns build up in the two different delay loops that have the corresponding recurrence times. Fig. 7 shows the response of the network to a concurrent double vowel, whose constituents are the synthetic vowels [ae] and [er] which respectively have fundamental periods of 100 and 125 Hz. The two constituent vowels have waveform patterns that repeat every 10 and 8 ms (top plots). The double vowel waveform was constructed by summing together the waveforms of the two constituent vowels. The waveform was then half-wave rectified and presented, as before, to all coincidence elements in the network.

One drawback of the simple 5% multiplicative rule of the last example is that it results in geometrically increasing signals, which over-emphasize waveform peaks. In this case a buildup rule that saturates more gracefully was chosen. Here 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 D_k, B is the adjustment/buildup rate factor (0.1), $S_{\text{direct}}(t)$ is the incoming direct input signal, and $S_{\text{loop}}(t)$ is the incoming signal circulating in the loop.

The behavior of the network is shown in the plot to the right of the waveforms. Within 2–3 periods, waveforms begin to build up in the two delay loops whose recurrence times equal the vowel periods, i.e. 10 ms (A) and 8 ms (B). The waveforms in the respective loops come to resemble the individual vowel constituents. This can be seen in the bottom plots (A and B). The top plots show the waveforms circulating in the two delay loops for the peristimulus time of 60–80 ms. In the 10 ms channel, peaks separated by 10 ms are seen; in the 8 ms channel, peaks are separated by 8 ms. The middle plots show the input to the network for the same time segment, and it can be seen that the waveforms in the two delay loops amplify different peaks in the composite double vowel input waveform. The bottom peaks show the constituent vowel waveforms for the same time period. Comparison with the loop waveforms indicates both common major and minor peaks.

This single-channel network demonstrates how multiple auditory objects with different repetition periods (i.e. fundamental periods, rhythms) can be segregated into different delay-paths. This is accomplished without any explicit estimation of the respective fundamentals and without the need to bind together particular channels or features to form each object. Building up and separating objects by temporal pattern coherences constitutes an extremely general and very powerful scene analysis strategy that potentially can be applied to any sensory system that has neural responses that are temporally correlated with the stimulus waveform.

4.3. Multichannel recurrent timing networks for separating and identifying concurrent (double) vowels

Most recently, recurrent timing networks have been scaled up to handle the multichannel temporal discharge patterns produced by a simulated auditory nerve array (Fig. 8). The network consisted of a simplified auditory nerve array front-end, and a full set of delay loops for each frequency channel.

The auditory nerve simulation incorporated bandpass filtering, half-wave rectification, low pass (synaptic) filtering, and rate compression. Twelve frequency channels were simulated with center frequencies spaced at equal logarithmic intervals from 125–4000 Hz. 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). Filters were fitted to approximate the rate responses of auditory nerve fibers (ANFs) as a function of tone frequency at a constant level of 60 dB SPL (Rose, Hind, Brugge & Anderson, 1971). Filtered signals were half-wave rectified and convolved with a square window low pass filter (i.e. a 200 μs moving average) that mimics the decline in phase-locking with frequency. An array of simulated peristimulus time histograms, called a PST neurogram, was thus generated. The PST pattern for each frequency channel was then fed to a set of 150 delay loops ranging from 0–15 ms recurrence time. The modified buildup rule that was described above was used to build up patterns in the loops.

A set of six concurrent, synthetic five-formant double vowels previously used in human psychophysics experiments (Assmann & Summerfield, 1989, 1990; Summerfield & Assmann, 1991) was presented to the whole network. Pairs of vowels [ae], [ah], [er], [ee], [oo], with same and different fundamental frequencies (0, 0.5, 1, 4 semitones apart, i.e. 0, 3, 6, and 24% difference in frequency) were used. Double vowels were 200 ms long. The responses to these double vowels were simulated, and the resulting PST discharge patterns were then presented to the recurrent timing nets. The outputs of the timing nets were then visualized and analyzed in several ways.

The response of the network to the double vowel [ae]–[er], with fundamentals of 100 and 106 Hz (one semitone apart) respectively, is shown in detail in Fig. 9(a)–(e). The first 50 ms of the input neurogram for this vowel is shown in Fig. 8. The network consists of 12 frequency channels with 150 delay loops per channel. The response is therefore described in three dimensions: frequency, delay loop, and time.

Panel A shows the average response of the network as a function of frequency and time, i.e. signals in all delay loops for a given frequency channel are summed together. Since the vowels excite overlapping frequency regions and the filters are relatively broad, there is a great deal of spectral overlap that provides few purely spectral cues for segregation.

Panel B shows the average response of the network as a function of loop delay and time. Here the signals from all loops across all frequencies having the same recurrence times have been summed together. The emergence of strong signals in loops with recurrence times of 10 and 9.4 ms, which correspond to the periods of the two vowels can be seen in the plot (gray arrows). Panel C shows the mean

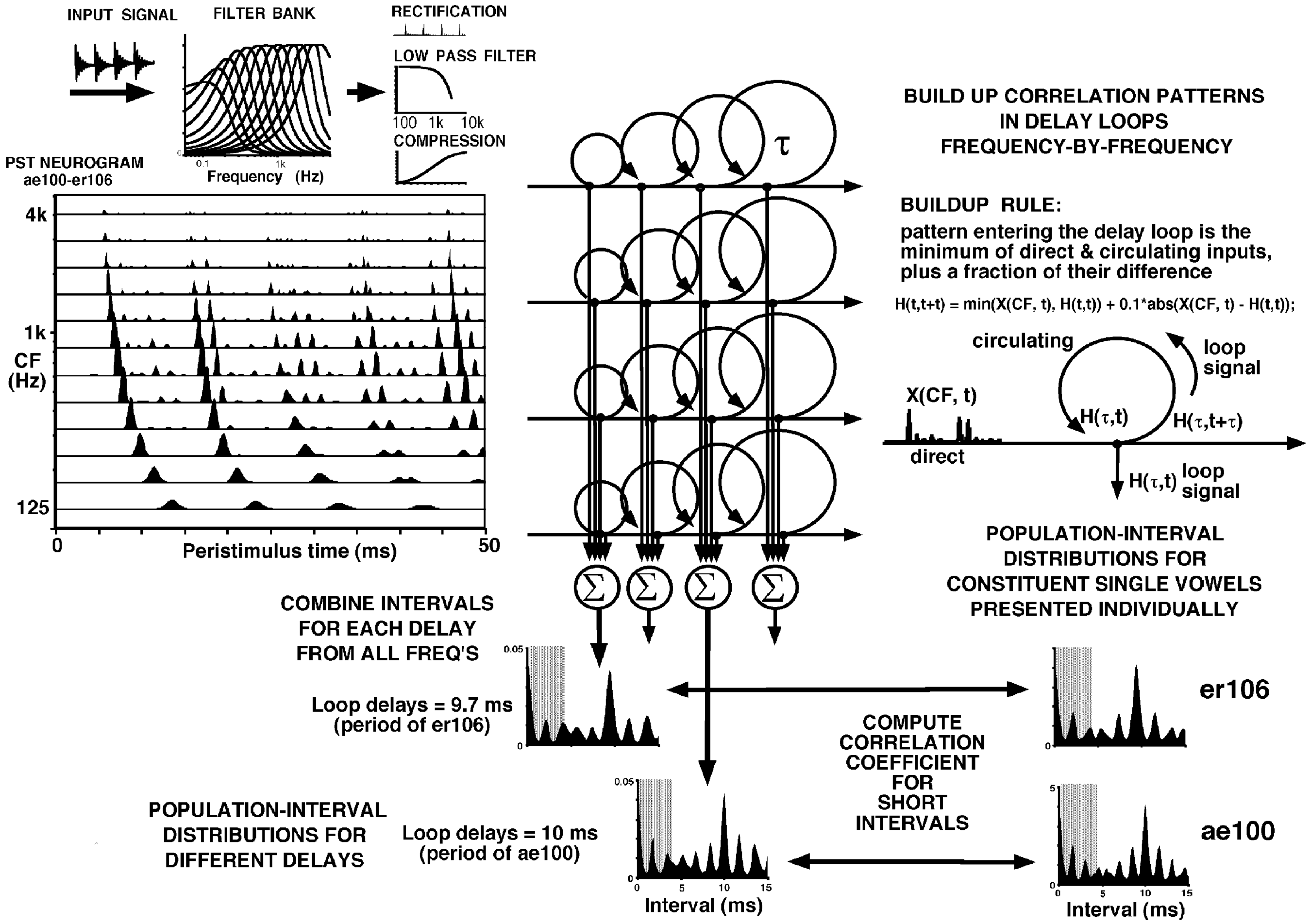

Fig. 8. Multichannel recurrent timing net. Schematic for multichannel, frequency-based recurrent timing net. Top left plots show stages in the auditory nerve simulation: band-pass filtering, half-wave rectification, low pass filtering, and rate compression. The simulation generates a PST neurogram (response to the double vowel [ae]–[er] is shown). Center, set of delay loops and coincidence elements for each frequency channel. Right, buildup rule. Bottom, comparison of loop-patterns with single vowel patterns using correlations between population–interval distributions.

signal value as a function of loop delay for three different time periods, i.e. a vertical cross section of the plot in panel B. In initial vowel periods, one peak is seen, followed by rapid separation in subsequent periods. The two peaks are at 9.4 ms (106 Hz fundamental of er) and 10 ms (100 Hz fundamental of ae). The response patterns for each 9.4 ms delay loop for each of the 12 center frequencies (loop neurogram) is shown in the middle plots, and the patterns for 10 ms delays are shown in the right hand plots. At the beginning, these are almost the same, but by 80–100 ms they are clearly different.

Panel D shows the separation of signals as a function of fundamental separation and buildup time. Fundamental separations of less than a semitone (6%) result in fused peaks, while separations of a semitone or more result in separation of signals into different delay channels. One wants to assess how similar the separated patterns are to those of the single vowels, and whether this similarity increases with their segregation into multiple delay channels. One means of doing this is to compile the population–interval distributions of the loop-neurograms and to compute their correlations with the population–interval distributions that are produced by the single vowels (Fig. 8, bottom). Panel E shows the population–interval distributions of the loop neurograms for the double vowel ae-er for different fundamental separations (0–4 semitones). As fundamental separations move from 0 to 4 semitones, correlations increase from 0.65 and 0.42 to 0.97. The greatest increase is between 0 and 1 semitone. Panel F shows the correlations of the signals in the dominant delay loops to those associated with their respective single vowels. In all cases, the general pattern of rapid improvement from 0–1 semitone followed by plateau from 1–4 semitones was observed. The high final correlations indicate that such networks effectively separate out the constituent vowels.

The network behaves qualitatively like human perception. When fundamentals are the same, the vowels are fused together, and their individual timbres are hard to hear out. When they are separated by a semitone ($\Delta F0 = 6\%$) or more, they are heard as two separate auditory objects, and are identified with somewhat higher accuracy. When human listeners are asked to identify the two constituents of double vowels,

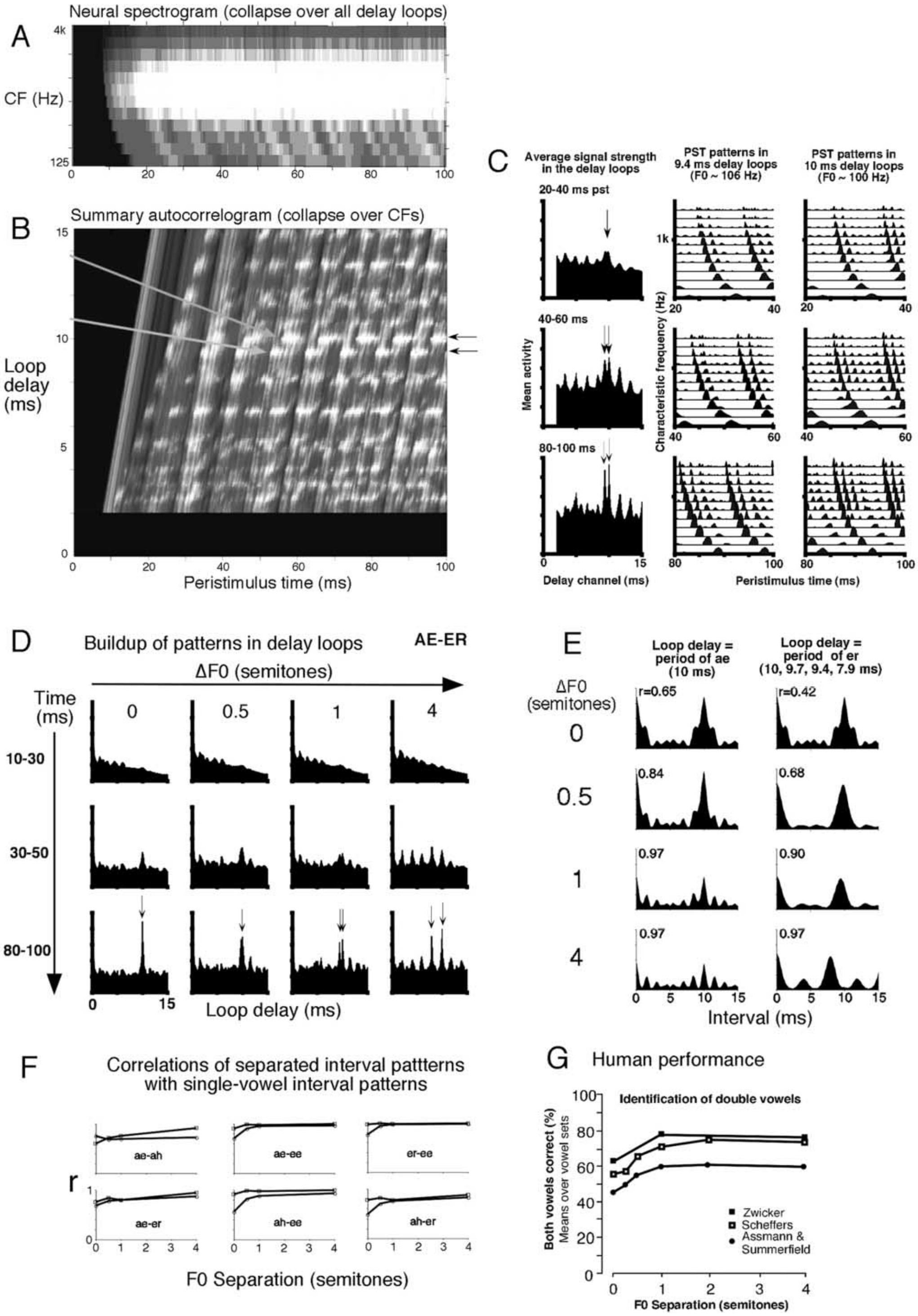

Fig. 9. Separation of concurrent double vowels using a multichannel recurrent timing net. Response to the double vowel ae–er, $F0(\text{ae}) = 100$ Hz, $F0(\text{er}) = 106$ Hz. (a) Plot of mean activation vs. time as a function of characteristic frequency (combining across loop delays). (b) Mean activation vs time as a function of loop delay (combining across frequencies). (c) Left, average signal strength in the delay loops as a function of loop delay and peristimulus time. Right plots, waveforms circulating in 10 and 9.4 ms delay loops at different peristimulus times. (d) Buildup of patterns in delay loops as a function of fundamental separation and peristimulus time. (e) All-order interval patterns in the delay loops. Increase in similarity between all-order interval patterns in the delay loops and those produced by their respective single vowels. (f) Correlations between loop interval patterns and their respective single vowels for all six double vowels, as a function of $F0$ separation. (g) Percent correct of double vowels whose constituents were correctly identified by human listeners as a function of $F0$ separation. Results of three studies (Assmann & Summerfield, 1989, 1990; Scheffers, 1983; Zwicker, 1984). Redrawn from Meddis and Hewitt (1992).

they correctly identify 45–65% of double vowels in sets correctly when the fundamentals are the same (panel G). They improve their identifications by 15–20% when the vowels are separated by a semitone or more. Thus far, a correlation-based decision rule has not been implemented that would allow more direct comparison between the network's error rates and those of human listeners. One possibility would be to examine all of the correlations between a given loop-pattern and prospective single vowel patterns, and to choose to identify the single vowel with the highest correlation. Similar correlation-based decision strategies were used successfully in the past to identify double vowels from neural ANF population–interval distributions (Cariani & Delgutte, 1993, 1994).

The model demonstrates that recurrent networks can be scaled up to handle multichannel input data, and that multiple auditory objects can be effectively separated using these techniques. Most existing strategies for separating sounds on the basis of fundamental frequency attempt to group frequency channels together by finding *F*0-related features in each channel, e.g. (Meddis & Hewitt, 1992). The present model demonstrates an alternate strategy for auditory object separation that uses no explicit feature detection (i.e. *F*0-detectors). Instead, the delay loops amplify temporal pattern-invariances that separate auditory objects on the basis of their temporal patterns. The network also demonstrates the buildup of auditory images in a manner not unlike 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 both architectures, object formation comes prior to analysis of auditory qualities (pitch, timbre) rather than being the result of such analyses. Finally, recurrent timing nets demonstrate how purely temporal representations and computations can effect separation and identification of auditory objects.

5. Future work

The simple timing nets presented here are certainly quite rudimentary, and there are many directions that timing net models and applications could pursue. The most obvious potential applications involve enhancement of periodic sounds in noise (voiced portions of speech in background noise) and separation of multiple periodic sounds, such as different speakers or musical instruments. We have also begun to examine possible applications of recurrent timing nets to the buildup of rhythmic expectations (Cariani, 1999a, 2001c).

Feedforward networks are useful in extracting which pitch- and timbre-related periodicities are common to their two inputs. This may be useful in determining speaker identity, which involves, among other factors, voice pitch comparisons (common fundamental frequency). Such mechanisms may also be useful in forming phonetic equivalence-classes based on vowel identity (timbre, largely irrespective of voice pitch).

At present timing nets function as broad heuristics for how the auditory system might process temporal patterns to form auditory objects and temporal expectations. Hypothetical grounding of these networks in specific neural substrates are beginning to be contemplated. The most obvious locus of feedforward timing nets would lie in the binaural cross-correlation operations situated in the nucleus of the medial superior olive (MSO). The idea for feedforward timing nets grew out of consideration of whether temporal patterns of binaural coincidences might be preserved in the output of the MSO, such that they could subserve perception of binaurally-created pitches (Akeroyd & Summerfield, 1999; Cariani, 1996, 2001a). More generally, Braitenberg (1961) proposed cortically-organized architectures for temporal processing in which horizontal fiber systems function as tapped delay lines and Purkinje/pyramidal cells function as coincidence detectors. There has been a running debate concerning the nature of cortical pyramidal cells, whether they are to be seen as rate integrators or coincidence detectors (Abeles, 1982). If pyramidal cells behave more like coincidence detectors, or that temporally-correlated activations of specific subsets of synapses are capable of initiating spikes, then fine timing issues rise to the fore in cortical structures (Abeles, Prut, Bergman & Vaadia, 1994), and notions of mass-statistical temporal processing in cortical coincidence arrays no longer appear so far fetched.

Recurrent pathways are the rule rather than the exception in the brain. Recurrent timing nets could potentially be realized via interactions between ascending and descending fiber systems at the level of colliculus and thalamus. Even at the level of the auditory thalamus, there exists enough phase-locked information to represent periodicities up to 2–3 kHz (de Ribaupierre, 1997), so that operations on interspike intervals at those stations are not completely out of the question (Cariani, 1999a). Thus far, there exist no satisfactory neurally-grounded accounts of how or where auditory images are formed and compared.

Adaptive resonance theory may provide a guide (Grossberg, 1988, 1995). Recurrent timing networks can be seen as temporal adaptive resonance networks in which patterns are temporally rather than spatially coded, and processing occurs in the time-domain. In both adaptive resonance and recurrent timing networks, the interplay of incoming sensory data and central circulating patterns results in bottom-up/top-down codeterminations. Although the timing nets presented here dynamically form patterns rather than using stored pattern archetypes to recognize incoming ones, central neural assemblies could emit temporal patterns that facilitate their buildup if they are present in incoming sensory data. Thus far, recurrent timing nets do not exploit mismatches between incoming patterns and network expectations as they do in adaptive resonance circuits. Nevertheless, 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. One would then have both coincidence and anticoincidence operations—correlation and cancellation (cf. Seneff's (1985, 1988) Generalized Synchrony Detector that computes the ratio of waveform sums and differences). 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, straightforward improvements can be made. Hebbian rules that operate on temporal correlations and anticorrelations, in the short-term as well as the long term can be incorporated. Perhaps the most exciting prospect is that delay loops could be formed on the fly even in randomly-connected nets by short-term facilitations borne by temporal correlations. The time structure of a incoming signal would dynamically organize central neural circuits so as to propagate and build up stable, reverberating patterns.

6. Conclusions

Neural timing nets are a class of neural networks that operate on temporally-structured spike patterns to produce other temporally-structured patterns. Neural timing nets implement time-domain operations on spike trains that are similar in style to analog signal processing.

A simple, feedforward coincidence array can operate on two sets of temporally-coded inputs in order to extract common periodicities underlying common pitches and timbres. Common pitch can thus be recognized independent of timbre, and common timbre can be recognized independent of pitch. This has the practical value of allowing one to extract common fundamentals (perceptually, pitches) even if there are no overlapping partials. Further, both operations can be realized using the same, simple mechanism that does not require explicit prior explicit estimation of either attribute.

Feedforward timing nets permit time patterns to be simply separated on the basis of differences in time-of-arrival. This provides an elegant mechanism for binaural separation and fusion.

Recurrent timing networks can build up periodic temporal patterns in their inputs and separate multiple auditory objects on the basis of differences in fundamental frequency. We have shown how such networks can build up and separate double vowels into their constituent waveforms. Recurrent timing nets implement alternative, global relational strategies for scene analysis that do not rely on binding together ensembles of local features into stable objects. Instead, such networks provide general-purpose pattern recognizers that form objects by fusing invariant temporal patterns in their inputs. Many other possible computational properties and uses of neural timing nets remain to be explored.

Acknowledgements

This work was supported by Research Grant DC03054 from the National Institute on Deafness and Other Communicative Disorders, US National Institutes of Health.

References

Abeles, M. (1982). Role of the cortical neuron: integrator or coincidence detector. *Israel Journal of Medical Sciences*, *18*, 83–92.

Abeles, M. (1990). *Corticonics*, Cambridge: Cambridge University Press.

Abeles, M., Prut, Y., Bergman, H., & Vaadia, E. (1994). Synchronization in neural transmission and its importance in information processing. In G. Buzsáki, R. Llinás, W. Singer, A. Berthoz & Y. Christen, *Temporal coding in the brain* (pp. 39–50). Berlin: Springer-Verlag.

Akeroyd, M., & Summerfield, Q. (1999). A fully temporal account of the perception of dichotic pitches. *Br. J. Audiol.*, *33* (2), 106–107 [abstract].

Assmann, P. F., & Summerfield, Q. (1989). Modeling the perception of concurrent vowels: Vowels with the same fundamental frequency. *J. Acoust. Soc. Am.*, *85*, 327–338.

Assmann, P. F., & Summerfield, Q. (1990). Modeling the perception of concurrent vowels: vowels with different fundamental frequencies. *J. Acoust. Soc. Am.*, *88*, 680–697.

Bialek, W., Rieke, F., van Stevenink, R. R., & de Ruyter, W. D. (1991). Reading a neural code. *Science*, *252*, 1854–1856.

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.

Braitenberg, V. (1967). Is the cerebellar cortex a biological clock in the millisecond range? *Prog. Brain Res.*, *25*, 334–346.

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

Cariani, P. (1996). Population–interval models for binaural periodicity pitches. *Soc. Neurosci. Abstr.*, *22* (1), 649.

Cariani, P. (1997a). Emergence of new signal-primitives in neural networks. *Intellectica*, *1997* (2), 95–143.

Cariani, P. (1997b). Temporal coding of sensory information. In J. M. Bower, *Computational neuroscience: Trends in research* (pp. 591–598). New York: Plenum.

Cariani, P. (1999a). Temporal coding of periodicity pitch in the auditory system: an overview. *Neural Plasticity*, *6* (4), 147–172.

Cariani, P. (1999b). Timing nets for rhythm perception (working paper). *Music and Timing Networks, Proceedings of the FWO Research Society on Foundations of Music Research, University of Ghent, October. 1999*, pp. 28–37.

Cariani, P. (2000). Auditory object formation through temporal coherence. *Assoc. Res. Otolaryn. Abstr.*, *23*, 102.

Cariani, P. (2001a). Neural timing nets for auditory computation. In S. Greenberg & M. Slaney, *Computational models of auditory function* (pp. 223–236). Amsterdam: IOS Press.

Cariani, P. (2001b). Temporal coding of sensory information in the brain. *Acoust. Sci. & Tech.*, *22* (2), 77–84.

Cariani, P. (2001c). Temporal codes, timing nets, and music perception. *Journal of New Music Research*, (in press).

Cariani, P., & Delgutte, B. (1993). Interspike interval distributions of auditory nerve fibers in response to concurrent vowels with same and different fundamental frequencies. *Assoc. Res. Otolaryngol. Abs.*, *16*, 373.

Cariani, P., & Delgutte, B. (1994). Transient changes in neural discharge patterns may enhance separation of concurrent vowels with different fundamental frequencies. *J. Acoust. Soc. Am.*, *95* (5(2)), 2842 [abstract].

Cariani, P. A., & Delgutte, B. (1996). Neural correlates of the pitch of complex tones: I. Pitch and pitch salience: II. Pitch shift, pitch ambiguity, phase-invariance, pitch circularity, and the dominance region for pitch. *J. Neurophysiology*, *76* (3), 1698–1734.

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).

Carney, T., Silverstein, D. A., & Klein, S. A. (1995). Vernier acuity during image rotation and translation: visual performance limits. *Vision Res*, *35* (14), 1951–1964.

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

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

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

Colburn, S. (1996). Computational models of binaural processing. In H. Hawkins & T. McMullin, *Auditory computation*, New York: Springer Verlag.

Colburn, S., & Durlach, N. I. (1978). Models of binaural interaction. In E. C. Carterette & M. P. Friedman, *Handbook of perception* (pp. 467–518). , Vol. IV. New York: Academic Press.

Culling, J. F., Summerfield, Q., & Marshall, D. H. (1998). Dichotic pitches as illusions of binaural masking release I: Huggins' pitch and the binaural edge pitch. *J. Acoust. Soc. Am.*, *103*, 3509–3526.

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

de Cheveigné, A. (1998). Cancellation model of pitch perception. *J. Acoust. Soc. Am.*, *103* (3), 1261–1271.

De Ribaupierre, F. (1997). Acoustical information processing in the auditory thalamus and cerebral cortex. In G. Ehret & R. Romand, *The central auditory system* (pp. 317–397). New York: Oxford University Press.

Di Lorenzo, P. M., & Hecht, G. S. (1993). Perceptual consequences of electrical stimulation in the gustatory system. *Behavioral Neuroscience*, *107*, 130–138.

Eggermont, J. J. (1990). *The correlative brain: Theory and experiment in neural interaction*, Berlin: Springer-Verlag Vol. 24.

Eggermont, J. J. (1993). Functional aspects of synchrony and correlation in the auditory nervous system. *Concepts in Neuroscience*, *4* (2), 105–129.

Emmers, R. (1981). *Pain: A spike-interval coded message in the brain*, New York: Raven Press.

Ghitza, O. (1988). Temporal non-place information in the auditory-nerve firing patterns as a front-end for speech recognition in a noisy environment. *Journal of Phonetics*, *16*, 109–123.

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

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

Grossberg, S. (1995). Neural dynamics of motion perception, recognition learning, and spatial attention. In R. F. Port & T. van Gelder, *Mind as motion: Explorations in the dynamics of cognition* (pp. 449–490). Cambridge: MIT Press.

Hirahara, T., Cariani, P., & Delgutte, B. (1996a). Representation of low-frequency vowel formants in the auditory nerve. *Assoc. Res. Otolaryng. Abstr.*, *19*, 80.

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

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

Johannesma, P., Aertsen, A., van den Boogaard, H., Eggermont, J., & Epping, W. (1986). From synchrony to harmony: ideas on the function of neural assemblies and on the interpretation of neural synchrony. In G. Palm & A. Aertsen, *Brain theory* (pp. 25–47). Berlin: Springer-Verlag.

Jones, M. R. (1976). Time, our lost dimension: toward a new theory of perception, attention, and memory. *Psychological Review*, *83* (5), 323–355.

Kauer, J. S. (1974). Response patterns of amphibian olfactory bulb neurones to odour stimulation. *J. Physiol.*, *243*, 695–715.

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.

Kozak, W. M., & Reitboeck, H. J. (1974). Color-dependent distribution of spikes in single optic tract fibers of the cat. *Vision Research*, *14*, 405–419.

Krishna, B. S., & Semple, M. N. (2000). Auditory temporal processing: responses to sinusoidally amplitude-modulated tones in the inferior colliculus. *J. Neurophysiol.*, *84*, 255–273.

Lange, F. H. (1967). *Correlation techniques*, Princeton: Van Nostrand (Johns, P. B., Trans.).

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

Laurent, G. (1999). A systems perspective on early olfactory coding. *Science*, *286* (5440), 723–728.

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

Licklider, J. C. R. (1959). Three auditory theories. In S. Koch, *Psychology: A study of a science. Study I. Conceptual and systematicSensory, perceptual, and physiological formulations* (pp. 41–144), Vol. I. New York: McGraw-Hill.

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

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

Lyon, R., & Shamma, S. (1995). Auditory representations of timbre and pitch. In H. Hawkins, T. McMullin, A. N. Popper & R. R. Fay, *Auditory computation* (pp. 221–270). New York: Springer Verlag.

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

Lyon, R. F. (1984). *Computational models of neural auditory processing*. Paper presented at the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), San Diego, March. 1984.

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

MacKay, D. M., & McCulloch, W. S. (1952). The limiting information capacity of a neuronal link. *Bull. Math. Biophys.*, 14.

Mead, C. (1989). *Analog VLSI and neural systems*, Reading, MA: Addison-Wesley.

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., & Hewitt, M. J. (1992). Modeling the perception of concurrent vowels with different fundamental frequencies. *J. Acoust. Soc. Am.*, *91*, 233–245.

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

Meyer-Eppler, W. (1953). Exhaustion methods of selecting signals from noisy backgrounds. In W. Jackson, *Communication theory* (pp. 183–193). London: Butterworths.

Michelson, A. (1992). Hearing and sound communication in smal animals: evolutionary adaptations to the laws of physics. In D. Webster, R. R. Fay & A. N. Popper, *The evolutionary biology of hearing* (pp. 49–60). New York: Springer-Verlag.

Miller, R. R., & Barnet, R. C. (1993). The role of time in elementary associations. *Current Directions in Psychological Science*, *2* (4), 106–111.

Moore, B. C. J. (1997). *Introduction to the psychology of hearing*, (4th ed). London: Academic Press.

Mountcastle, V. (1967). The problem of sensing and the neural coding of sensory events. In G. C. Quarton, T. Melnechuk & F. O. Schmitt, *The neurosciences: A study program*, New York: Rockefeller University Press.

Pabst, M., Reitboeck, H. J., & Eckhorn, R. (1989). A model of preattentive texture region definition based on texture analysis. In R. M. J. Cotterill, *Models of brain function* (pp. 137–150). Cambridge: Cambridge University Press.

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, *The auditory processing of speech* (pp. 115–124). Berlin: Mouton de Gruyter.

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.

Rees, A., & Møller, A. R. (1987). Stimulus properties influencing the responses of inferior colliculus neurons to amplitude-modulated sounds. *Hearing Res.*, *27*, 129–143.

Reichardt, W. (1961). Autocorrelation, a principle for the evaluation of sensory information by the central nervous system. In W. A. Rosenblith, *Sensory communication* (pp. 303–317). New York: MIT Press/John Wiley.

Reinagel, P., & Reid, C. (2000). Temporal coding of visual information in the thalamus. *J. Neurosci.*, *20* (14), 5392–5400.

Reitboeck, H. J. (1989). Neural mechanisms of pattern recognition. In J. S. Lund, *Sensory processing in the mammalian brain* (pp. 307–330). Oxford: Oxford University Press.

Reitboeck, H. J., Pabst, M., & Eckhorn, R. (1988). Texture description in the time domain. In R. M. J. Cotterill, *Computer simulation in brain science*, Cambridge, UK: Cambridge University Press.

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

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

Scheffers, M. T. M. (1983). *Sifting Vowels: Auditory Pitch Analysis and Sound Segregation*. Unpublished PhD, Rijksuniversiteit te Groningen, The Netherlands.

Sejnowski, T. J. (1999). Neural pulse coding. In W. Maass & C. M. Bishop, *Pulsed neural networks* (pp. xiii–xxvi). Cambridge, MA: MIT Press.

Seneff, S. (1985). *Pitch and Spectral Analysis of Speech Based on an Auditory Synchrony Model*. Unpublished PhD, MIT.

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

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.

Singer, W. (1995). Time as coding space in neocortical processing. In G. Buzsáki, R. Llinás, W. Singer, A. Berthoz & Y Christen, *Temporal coding in the brain* (pp. 51–80). Berlin: Springer-Verlag.

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

Summerfield, Q., & Assmann, P. F. (1991). Perception of concurrent vowels: effects of harmonic misalignment and pitch-period asynchrony. *J. Acoust. Soc. Am.*, *89* (3), 1364–1377.

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

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

Uttal, W. R. (1973). *The psychobiology of sensory coding*, New York: Harper and Row.

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

von Bekesy, G. (1967). *Sensory inhibition*, Princeton: Princeton University Press.

Wang, D. L. (1995). An oscillatory correlation theory of temporal pattern segmentation. In E. Covey, H. L. Hawkins & R. F. Port, *Neural representation of temporal patterns* (pp. 53–76). New York: Plenum Press.

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

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

Young, R. A. (1977). Some observations on temporal coding of color vision: psychophysical results. *Vision Research*, *17*, 957–965.

Young, E. D., & Sachs, M. B. (1979). Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory nerve fibers. *J. Acoust. Soc. Am.*, *66* (5), 1381–1403.

Zwicker, U. T. (1984). Auditory recognition of diotic and dichotic vowel pairs. *Speech Commun.*, *3*, 265–277.