

## NEURAL TIMING NETS FOR AUDITORY COMPUTATION

Peter Cariani

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

### 1. Introduction: Temporal Coding of Auditory Qualities

Pitch, timbre, and rhythm are basic auditory qualities that are fundamental to the perception of speech, music and environmental sounds. These perceptual qualities have much in common:

- (1) they are very precise (subtle discriminations can be made),
- (2) they are largely invariant in the face of large changes in stimulus intensity, location in auditory space, and background noise levels, and
- (3) they are apprehended by a wide variety of animals.

A central goal for auditory physiology has always been to understand the nature of the neural codes, representations and processing architectures that subservise these auditory form-percepts. Many auditory physiologists and psychoacousticians have recognized the pervasive parallels that exist between auditory percepts on one hand and the temporal discharge patterns of auditory neurons on the other. On many levels, the properties of neural representations based on the stimulus-locked character of neural discharge patterns (spike timings, synchronicity, interspike intervals) mirror those common properties listed above.

The strongest candidate neural codes for pitch at the level of the auditory nerve and brainstem are those based on all-order, interspike interval distributions of populations of auditory neurons ("population-interval distributions"). Historically a diverse array of models and simulations has pointed to the use of interspike-interval information by the auditory system in explaining the various pitches that are heard [33][34][37][39][43][56] as well as the precision with which they can be discriminated [21][43][55]. While a large number of neurophysiological studies of the auditory nerve have examined the interspike-interval correlates of pitch perception and frequency discrimination, it has only been relatively recently that population-interval distributions have been estimated from auditory-nerve data [10][11][45]. In our own investigations [10][11], we found that features of population-interval distributions estimated from observed responses of 50-100 single auditory-nerve fibers of Dial-anesthetized cats closely parallel those of human pitch perception [10][11]. With very few exceptions, the most frequent interval in the auditory nerve at any given time corresponds to the pitch that is heard. Many complex pitch-related phenomena are readily explained in terms of these population-interval distributions: the pitch of the missing fundamental, pitch equivalence, relative phase and level invariance, non-spectral pitch, pitch shift of inharmonic tones and the dominance region.

We have also observed empirically that patterns of major and minor peaks in population-representations resemble those of their respective stimulus autocorrelation functions [9]. In retrospect, it has become apparent that this similarity is a general consequence of the phase-locking of neural discharges. Because phase-locked responses are found in many other sensory systems, such as vision, mechanoreception, and electroreception, this finding has broad

implications outside of the auditory system [4][7][47]. To the extent that a receptor system produces neural discharges whose timings are highly correlated with stimulus time structure, distributions of all-order interspike intervals resemble the stimulus autocorrelation function. In the auditory system, by virtue of the phase-locking abilities of auditory neurons, population-interval distributions provide very general autocorrelation-like representations for stimulus periodicities up to the limits of phase-locking.

Population-interval distributions representations are also capable of representing the timbres of stationary sounds, such as vowel quality [4][9][25][37][40][45]. Timbre is associated with shapes of spectral envelopes, which manifest themselves in autocorrelation functions as patterns of minor peaks (Figure 3). To the degree that each stimulus component produces phase-locked discharges, it contributes its time structure to the population interval distribution. Consequently, in the auditory nerve, different vowels, with different sets of dominant frequency components, produce population-interval distributions with characteristic patterns of short intervals that reflect their respective formant structures. Changes in these population-interval patterns closely follow vowel-identification boundaries [25].

Population-interval distributions thus appear to be capable of subserving a wide variety of auditory qualities associated with pitch and timbre. These strong psycho-neural correspondences beget questions of whether the central auditory system, in fact, utilizes this interval-based information, and if it does use such information, how it uses it. Related to these questions are still others that concern the fate of neural timing information as one ascends the auditory pathway. Is the neural timing information that is so precise and robust and in such abundance at the level of the auditory nerve converted to across-neuron patterns of activation in higher, central auditory stations? Or is the temporal structure preserved in some way, perhaps in less synchronous and more spatially distributed form than is found in lower stations? If temporal information is in fact available in central auditory stations at the level of the midbrain, thalamus, and/or cortex, what kinds of neural processing architectures would be needed to make use of it?

This paper explores some possible means by which neural networks might analyze distributed, population-based temporal representations of auditory qualities. For the most part we will put aside for the present questions of where these neural networks might be concretely located, in favor of more functionally oriented ones devoted to exploring their potential information-processing capabilities. Whether these kinds of neural computations are in fact carried out in central auditory structures are empirical questions that can only be answered through directed neurophysiological experiments. While a detailed understanding of how the auditory portion of the brain works as an information-processing system remain our ultimate goal, we can only direct our neurophysiological lenses effectively if we already have some strong ideas of the kinds of neural computational mechanisms that might be possible.

## **2. Time-to-Place Conversions**

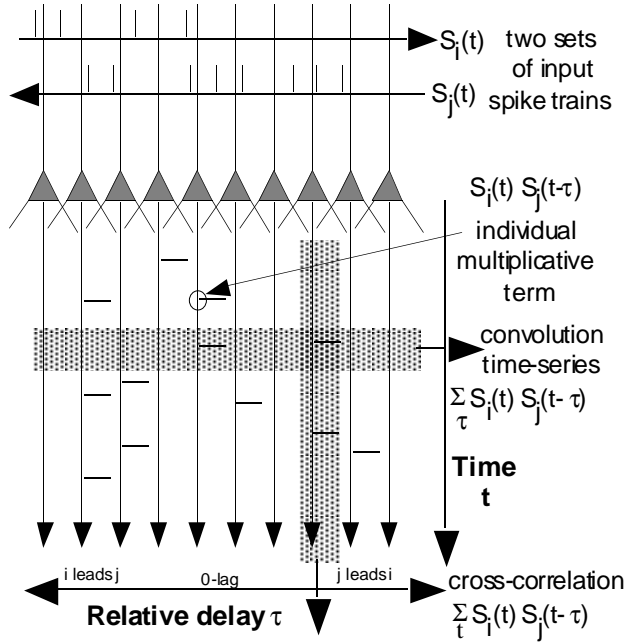
In the past virtually all of the temporal theories of hearing have assumed that the temporal information found at the level of the auditory nerve is converted to spatial patterns of activation somewhere higher in the auditory pathway. Many of the first neural networks that were proposed for auditory computation, such as the Jeffress model for auditory localization [27] and the Licklider duplex model for pitch perception [33], were time-delay neural networks whose purpose was to carry out this conversion. It was generally assumed that the outputs of such networks would then be analyzed via traditional, channel-coded connectionist

networks in more central stations. For example, Licklider's time-delay architectures [33][34] converted temporal input patterns to spatialized autocorrelation profiles by means of delay lines and coincidence detectors. However, sharply tuned autocorrelator-like periodicity detectors have yet to be found in the auditory pathway. Likewise, neurophysiological investigations in the auditory cortex have failed to find other kinds of simple pitch-detection units [54]. The most promising evidence for a time-to-place transformation has involved the modulation-tuning properties of central auditory neurons [32][53]. However, modulation-tuning tends to be relatively coarse, and to weaken at higher levels and in background noise [48][49]. Moreover, as one ascends the auditory pathway to auditory midbrain, thalamus, and cortex, best modulation frequencies (BMFs) generally decline, with progressively fewer BMFs covering the periodicity pitch range (50-500 Hz). This shift towards lower BMFs parallels declines in average discharge rates and synchronization indices that are seen. Finally, modulation-based representations, like first-order interval detectors, sometimes diverge from the autocorrelation-like behavior that characterizes pitch judgments (e.g. de Boer's rule for pitch shifts of inharmonic complex tones).

### 3. Neural Timing Networks: Time-Time Comparisons

A second possible strategy for representing and analyzing auditory forms is to retain temporal information in one form or another, and to perform comparisons between different time patterns by observing their interactions. For example, one can detect extremely subtle differences in frequency by binaural comparisons in which one listens for the presence of binaural beats. A major question for such an approach concerns the availability of temporal information to be analyzed. Unfortunately, the existence limits of neural timing information in the auditory pathway are still not well established. Pitch-related temporal patterns are omnipresent in the auditory nerve and cochlear nucleus [5][51] and are still quite evident in the auditory midbrain [22][32]. Although neural interspike interval information present in single units thins out dramatically as one proceeds from brainstem to thalamus to cortex, it is nevertheless possible that the requisite timing information to support central time codes for pitch and timbre exists in thalamocortical loops. Roughly half of all units encountered in lightly-anesthetized auditory thalamus show significant phase-locking (synchronization index  $> 0.3$ ) to pure tones of 250-500 Hz, while roughly 10% phase-lock to 1-2 kHz tones [20]. Response periodicities of several hundred Hz are observed in unanesthetized primary auditory cortex [20][57]. To the extent that interspike interval information exists in many of these stations, it remains precise, robust, and faithful to the autocorrelation-like behavior of pitch. It is important to remember that the timing information present in the auditory nerve far exceeds that required for human frequency discrimination [21][55]. Accordingly, only a small fraction of the timing information available at the auditory nerve need be faithfully transmitted and preserved for central auditory analyzers in order to realize the perceptual capabilities that are observed for the organism as a whole.

If the interval-based information is indeed available in central auditory stations, what kinds of neural networks are required for its analysis? 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 (time-to-time mappings). Their closest precursors are simple functional models of neural computation for which fine time structure is of primary importance [1][3][12][27][34][35][36][38][46][50][58]. Some of these precursors were themselves inspired by the functional anatomy of cortical structures [3][50][58].

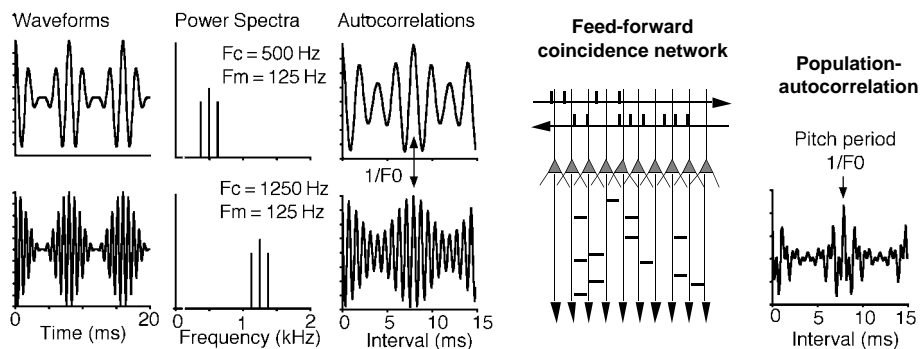


**Figure 1** Simple feedforward timing net consisting of an array of coincidence detectors and two sets of tapped delay lines through which input signals  $S_i$  and  $S_j$  arrive.

#### 4. Simple Feedforward Timing Nets

Consider an array of coincidence detectors that have inputs from two sets of tapped delay lines arranged in anti-parallel orientation (Figure 1). The configuration is reminiscent of both the Jeffress binaural localization model [27] and the Braitenberg cerebellar timing model [3]. 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. Thus, all relative delays are realized up to the conduction time across the array. Each coincidence detector requires nearly simultaneous arrival of a spike in both lines in order to fire. Consequently, each spike in the output of the coincidence array represents the joint occurrence of spike arrivals in the two inputs (or the multiplication of binary inputs,  $S_i(t) * S_j(t-t)$ ). A further consequence is that each interspike interval or higher-order spike arrival pattern appearing in a given output channel must also be present in each of the two inputs. Thus the array functions as a temporal sieve, passing those temporal patterns that are common to both sets of inputs. Several basic computations can be carried out. First, the cross-correlation function of the two inputs can be computed by counting the number of spikes in each output channel as a function of relative delay. Their convolution can be computed by summing across relative delay channels for each time step. Similarly, the summary- or population-autocorrelation of the outputs can be computed by summing the autocorrelations of each of the output channels.

The conduction time across the array implements a temporal contiguity window: those inputs that arrive within this time window interact, while those arriving at different times do not. All intervals from each set of inputs that arrive within the temporal contiguity window cross their counterparts, such that if one input has  $M$  intervals of duration,  $t_0$ , and the other has  $N$  such intervals, then  $M * N t_0$  intervals will appear in the outputs. Within the temporal



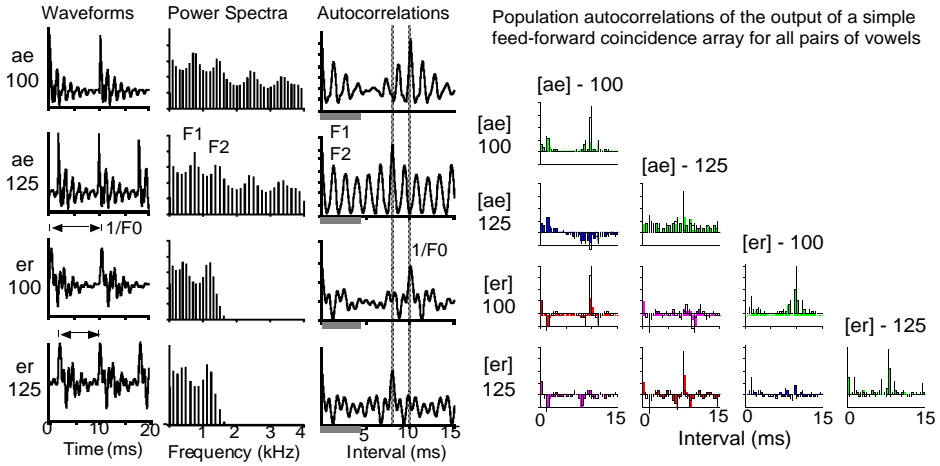
**Figure 2** Effect of passing two signals through the coincidence array. The stimuli are two AM tones with different carriers ( $f_c = 500$  Hz, 1250 Hz) but the same modulation frequency ( $f_m = 125$  Hz). The AM tones have no harmonics in common, but they produce a common low pitch at their “missing fundamental” ( $f_0 = f_m = 125$  Hz, dotted lines). Right: Population autocorrelation of the output of the coincidence array.

contiguity constraints, the coincidence array therefore performs a multiplication of the autocorrelations of its inputs.

The population-autocorrelation output of such a coincidence array is largely phase-insensitive. Because all of the intervals in the two input lines arriving within the time window cross their counterparts somewhere in the array, the short-term temporal ordering of the intervals within each incoming pulse train signal has little effect on the population-autocorrelation of the output. This behavior is qualitatively similar to the phase-insensitive character of auditory form perception: in general, we have great difficulty distinguishing pitches or timbres of complex tones that differ only in their phase spectra. Temporal contiguity constraints also exist in pitch and timbre perception. Pitches associated with the missing fundamental can be evoked for sets of harmonics that are presented successively, but disappear when brief periods ( $> 10$  ms) of silence are inserted between them [24]. Similarly, two single-formant vowels do not produce a two-formant vowel quality unless the waveforms corresponding to the two formants arrive within a similarly brief time window [13][14]. Provided that their waveforms overlap in time within this window and have the same fundamental, one cannot generally distinguish between combinations of single formant vowels with different relative delays among the vowels. The phase-insensitive nature of this coincidence array means that the mechanism can accommodate a good deal of asynchronous, temporal shifting among its inputs.

## 5. Recognition of Common Pitch Irrespective of Timbre

Coincidence arrays can extract those periodicities common to their inputs, even if their inputs have no harmonics in common. This is useful for the recognition of common pitches irrespective of differences in timbre (e.g. two different musical instruments playing the same note). As an example, two amplitude modulated (AM) tones were passed through the coincidence array (Figure 2). The fundamental frequency ( $f_0$ ) of an AM tone is equal to its modulation frequency ( $f_m$ ). AM tones produce strong pitches at their fundamental frequencies, despite the lack of any stimulus energy at that frequency (i.e., AM tones produce pitches at various “missing fundamentals”). For this example, the fundamental frequencies of the two signals were both set to 125 Hz, such that the signals produce the same low pitch at that fre-



**Figure 3** Left: Waveforms, power spectra, and autocorrelation functions for four vowels. The vowel set consists of combinations of two different fundamental frequencies ( $f_0 = 100, 125$  Hz) and two formant structures. Horizontal arrows above waveforms and vertical lines in autocorrelations indicate fundamental periods ( $1/f_0 = 8, 10$  ms), which correspond to voice pitch periods. Shaded bars indicate periodicities associated with formant structures that give rise to differences in vowel quality (timbre). Right: Population autocorrelations of the output of the coincidence array for all vowel pairs.

quency. Despite their common fundamental, the two signals have different carrier frequencies ( $f_c = 500$  Hz vs. 1250 Hz) and therefore have different spectral energy distributions. Such signals would produce different timbres. When the two signals are passed through the array, the resulting population autocorrelation is dominated by intervals at the common fundamental period,  $1/f_0 = 8$  ms. The array thus extracts those periodicities that are common to the two signals, and the form of those common temporal patterns appears directly in its output.

## 6. Recognition of Common Timbre Irrespective of Pitch

Coincidence nets can also extract common periodicities that are associated with different timbres or vowel qualities. This is useful for recognizing common timbres irrespective of differences in pitch (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 ( $f_0$ s) and two sets of formants ( $F_1, F_2, F_3, F_4, F_5$ ) were constructed (Figure 3). These signals correspond to the vowels [ae] (as in “hat”) and [ɜː] (as in “herd”). Their waveforms, power spectra, and autocorrelation functions are shown in Figure 3 (left). Each vowel evokes a “voice pitch” at its fundamental. Fundamental frequencies ( $f_0$ ) correspond to spacings between adjacent harmonics in the power spectra; fundamental periods ( $1/f_0$ ) correspond to major peaks in the respective autocorrelation functions. Each vowel also has a characteristic tonal quality (“timbre”) which determines whether it will be recognized as an [ae] or an [ɜː] (or some other vowel). The general shape of the power spectrum (spectral envelope) largely determines the timbre of a stationary sound; the spectral envelope, in turn, is largely shaped by positions and magnitudes of spectral peaks (formants). Different combinations of formants produce characteristic patterns of short time intervals in

the autocorrelation functions. Similar characteristic patterns corresponding to the fundamental and to formant combinations are observed in population-interval distributions at the level of the auditory nerve [4][8][9][37][45].

All combinations of the four waveforms were passed pairwise through the coincidence net (Figure 3, right panel). Population autocorrelations produced by vowels paired with themselves are equivalent to their own autocorrelations squared. Those vowel pairs that had common fundamental frequencies and similar voice pitches produced large peaks at their common fundamental periods. Those vowel pairs that had common formant structures (common vowel quality or timbre) produced common patterns of short intervals that correspond to their respective formant structures. Those vowel pairs that had neither common fundamental frequency nor common formant structure (different voice pitches and timbres) produced only small peaks associated with overlapping subharmonics.

Thus, 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. This permits a common pitch to be recognized independent of timbre, and a common timbre to be recognized independent of pitch. Further, both operations can be realized using the same, simple mechanism that operates on the interspike-interval statistics of an entire ensemble of neural elements.

## 7. Binaurally Created Pitches

The feedforward coincidence operations outlined above require the two sets of inputs to be simultaneously present in the network in order to effect pitch and timbral comparisons. The most obvious locations in the auditory system where one has simultaneous phase-locked inputs, tapped-delay lines, and arrays of coincidence detectors are structures in the auditory brainstem that receive binaural inputs. Low pitches and rhythmic binaural beats can be created by binaural interactions within these structures [15]. Historically, the existence of “binaurally-created pitches” was used to argue against temporal models for pitch that required interaction of neighboring harmonics within the same cochlea (e.g. Schouten’s theory of ‘residue’ pitch [18]). Houtsma and Goldstein [26] showed that binaural combination of two harmonics of a common fundamental could give rise to a binaurally created pitch at the missing fundamental. The existence of these pitches was explained in terms of a spectral-pattern analysis of harmonic structure in a “central spectrum” representation. The feedforward operation outlined above provides a temporal account of the generation of such pitches. Here the two sets of inputs to the coincidence array come from the auditory pathways originating in each ear. As with the two AM tones illustrated above, when two harmonically related pure tones are passed through such a coincidence network, the population autocorrelation function of the output produces a maximum at their common fundamental period. A similar result is obtained if the two harmonics are band-passed filtered, half-wave rectified, and the output of each channel is passed through a similar cross-correlation array [6]. The time-structure of the respective tones are impressed on swaths of frequency channels that overlap and these beat at the fundamental frequency. In those channels, binaural coincidence detectors consequently produce many intervals at the “missing” fundamental period. According to a general temporal autocorrelation theory of pitch, such a population-interval pattern would then be interpreted by central analyzers, much in the same way as monaural pitches, with the result that a binaural interaction pitch at the missing fundamental

should be heard. These observations notwithstanding, there are other temporal mechanisms, such as a simple central addition of the monaural population-interval distributions, that would also produce these pitches.

There are also other kinds of pitches that are created through binaural phase differences [2][16] that create troughs in the population autocorrelations of binaural cross-correlation arrays. These troughs correspond to the pitch periods that are heard. Such pitches therefore require cancellation or anti-correlation operations rather than simple coincidence operations [17][19]. Such operations could be incorporated into feed-forward timing nets by adding anti-coincidence detectors that produce output pulses when there is an incoming pulse in only one of the two input lines (an XOR operation). Once both coincidence and anti-coincidences are computed, timing networks attain the means of computing both temporal similarities and differences present in their inputs.

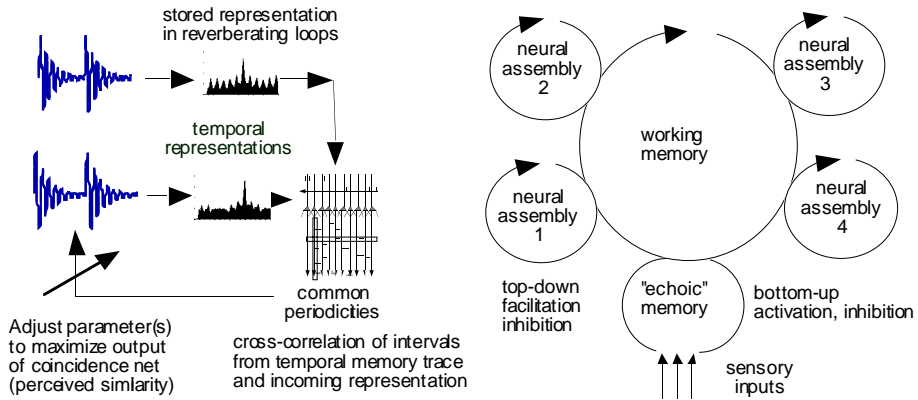
## **8. Simple, Recurrent Timing Nets**

The simple feed-forward networks outlined above carry out comparisons between inputs that are simultaneously presented. In order to perform delayed matching tasks, such networks would require some mechanism for maintaining a working memory representation of what came before. Perhaps the simplest means of storing time patterns, either in the form of post-stimulus time patterns or interval statistics, is to let the signals themselves circulate in recurrent sets of delay lines (Figure 4). A reverberating memory is thereby created in which the signal itself serves as its own temporal memory trace. Incoming time patterns can then be compared with those that are circulating using the kinds of feed-forward correlational operations outlined above. Matching of pitches or timbres in such a system then involves maximizing the correlation between the stored temporal pattern and the incoming one.

In such a system, recognition operations can be carried out if there are central neural assemblies that can produce temporal patterns that are characteristic of the objects to be recognized (e.g., interval distributions characteristic of particular vowels). Neural responses consistent with this notion have been observed in some neurophysiological conditioning studies [29][44][58], where stimulus-related temporal patterns are “assimilated” by individual neurons and “readout” at different times. If the outputs of an ensemble of such assemblies are cross-correlated with incoming temporal patterns and fed back into the loop, then those incoming patterns that resonate most strongly with those produced by neural assemblies will build-up the fastest. Strongly activated central temporal pattern templates can steer the build up of circulating patterns, such that the resulting resonances resemble the intersection of the incoming pattern with the stored templates, thereby creating “perceptual magnet effects.”

A considerable body of psychological evidence exists for mechanisms that build-up, store, and read-out temporal expectations. Studies of conditioning [28][42][58], music perception [30][31], and rhythm production [52] suggest that temporal relationships are explicitly encoded in memory, and that these relationships create sets of temporal expectancies. Recurrent timing nets implement reverberating memories [58] that can dynamically create short-term expectancies and build up temporal patterns that recur over time.

Perhaps the simplest example of a reverberating memory is the recurrent timing net shown in Figure 5. This network cross-correlates incoming time patterns with previous, circulating ones in order to build up those temporal patterns that recur. The network consists of an array of coincidence detectors which all receive the same external signals. Each coincidence detector has an associated delay loop with a different recurrence time. Coincidence



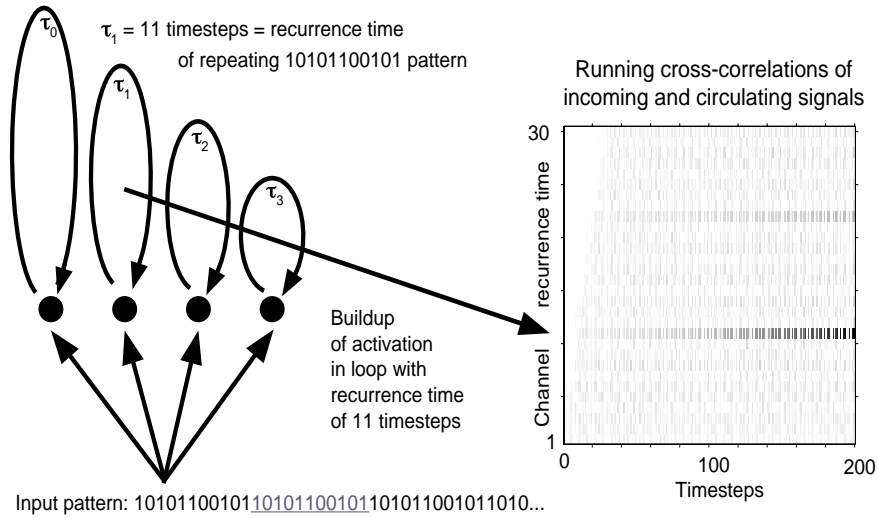
**Figure 4** Temporal memory traces, matching tasks, and the build-up of perceptual forms

detectors temporally cross-correlate incoming signals with those that are arriving via the delay loop. As a first step, pulse trains with repeated, randomly selected pulse patterns (e.g. 100101011-100101011-100101011...) are presented to the network. At each time step, the incoming pulse train is multiplied by the circulating pulse train arriving in each respective delay loop and the resultant signal is fed back into the loop. In the absence of pulses arriving through the delay loop, the incoming pulse train is fed into the loop. If there are coincident pulses from both inputs, the amplitude of the output pulses that are propagated through the loop is increased by 5%. Coincidences, therefore, build up the strength of the circulating pattern.

In such a network periodic pulse patterns invariably build up fastest in the delay loop whose recurrence time matches their repetition time. In their respective loops, rhythmic input patterns create temporal expectancies (when pulses traveling through the loop arrive back at the coincidence detector that generated them) that are reinforced when they are satisfied. Thus, recurrent time patterns are repeatedly correlated with themselves to build up to detection thresholds. In effect, the recurrent cross-correlation loops dynamically create matched filters from repeating temporal patterns in the stimulus. Thus, temporal-pattern invariances are enhanced relative to aperiodic transient activity, such as noise. Similar strategies for periodicity detection were explored in the 1950s [41].

More elaborate recurrent timing nets would also incorporate anticoincidence elements that compute the difference between expectation and the incoming signal. Once both correlation and anti-correlation operations are in place, these networks begin to resemble simplified, time-domain versions of adaptive resonance networks [23]. In place of spatialized input patterns and spatial pattern correlation operations for comparing them, timing nets utilize temporal input patterns, delay lines and coincidence detectors to do the comparisons in the time domain. Temporal correlation and anti-correlation take the place of excitation and inhibition. Both kinds of networks utilize recurrent bottom-up, top-down interactions to build up resonant patterns of activity. When inputs confirm top-down expectations, those expectations are reinforced; when inputs diverge from expectations, their differences form new expectation patterns that can then subsequently be built up.

These simple recurrent timing networks can also separate multiple time patterns with different repetition periods. When two repeating pulse patterns, each with its own repetition period, are summed and presented to the network, the two patterns invariably build up in the two different delay paths that have the corresponding recurrence times. These recurrent timing architectures were inspired by rhythm perception and production (e.g., [30][31][52]),

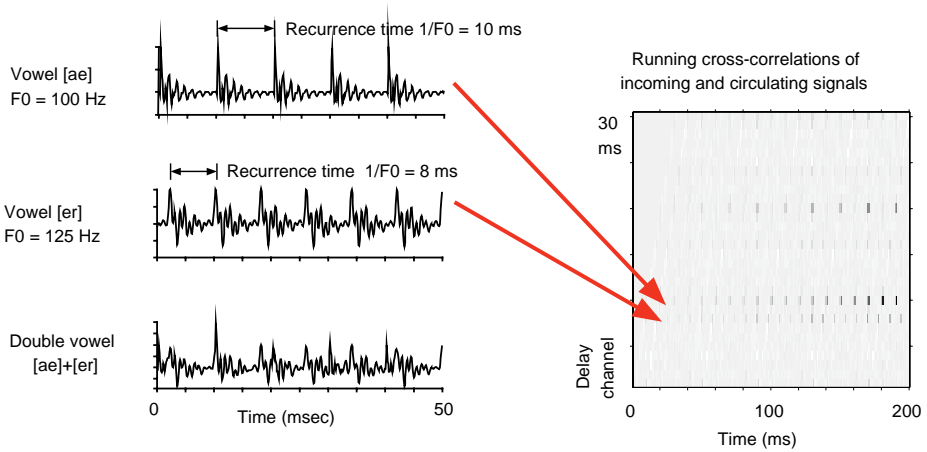


**Figure 5** Behavior of a simple recurrent timing net for periodic pulse-train patterns.

and phase-sensitive processes in auditory temporal integration [46]. While they were conceived to operate over longer time windows associated with these phenomena ( $> 30$  ms), many parallels exist between rhythm and pitch, such that these general processing strategies appear to be potentially applicable to pitch-related separations as well.

Two vowels with different fundamental frequencies ( $f_0 = 100$  Hz, 125 Hz) were summed together and presented to the recurrent network (Figure 6). Each period of the two vowels has its own invariant waveform pattern. The internal relations within the vowel periods of each waveform remain constant from period to period, whereas the relation between the two vowel-period waveforms change over time - the vowel periods precess relative to each other, creating “pitch period asynchronies.” Similar precessions and perceptual separations occur when an individual frequency component of a harmonic complex is mistuned. As with pairs of repeating pulse patterns, the two vowels build up their respective waveform patterns in the corresponding delay loops. (A potential problem with this multiplicative [vs. additive] buildup is that successive multiplications alter relative amplitudes of waveform peaks, although zero-crossings remain intact.) Thus, multiple auditory objects with different repetition periods (i.e. fundamentals, rhythms) can be segregated into different delay paths. Fusion is the consequence of recurrent, invariant temporal relations, while segregation is the consequence of changing temporal relations (precession of vocalic periods relative to each other).

Segregation by temporal pattern invariance constitutes an extremely general strategy for the formation and separation of perceptual objects. Traditional strategies for scene analysis are based on channel selection. First, a local feature analysis is carried out on incoming sensory patterns and an attempt is made to select subsets of feature channels that should be grouped together or separated to form different objects. For concurrent vowels, this has meant detecting which frequency channels share common  $f_0$ -related modulations and grouping them together (e.g. [40]). The correlational strategy proposed here instead groups patterns of spikes rather than patterns of channels. Here no explicit feature detection is required prior to the formation of auditory objects – the temporal patterns build themselves up and sort themselves out in their respective delay channels.



**Figure 6** Separation of two auditory objects, with differential fundamental frequencies, in a simple recurrent net.

Both feedforward and recurrent timing networks share a number of general functional properties that are highly desirable in the context of neural computation in the brain:

- 1) no highly tuned delay lines, periodicity detectors, or clocks are needed because no explicit time measurements are made,
- 2) representational precision resides in spike timings instead of in neural activation profiles,
- 3) harmonic relations implicit in time intervals are preserved (e.g. octave similarities, characteristic musical interval patterns), and
- 4) population-wide operations that make use of all neural responses, even weak ones, obviate the need to select relevant subpopulations for analysis.

Population-based temporal representations permit information from whole neural populations to be exported *en masse* to other regions. Coincidence networks permit comparisons between activity patterns of neural populations without the necessity of precise point-to-point mappings between them and/or highly regulated synaptic weightings. These properties may greatly simplify the coordination of information processing in large numbers of semi-independent, largely asynchronous populations of neurons.

How such computational strategies might scale up for large numbers of inputs, delay paths and coincidence elements remains to be explored. Simultaneous arrival of incoming pulses in three sets of inputs as a requirement for coincidence leads to higher-order, triple-correlation functions [59] that carry temporal sequence and phase information. Recurrent delay loops can be implemented by multisynaptic pathways, provided that the build-up of jitter can be constrained through general connectivity rules (e.g., fan-in/fan-out factors) or through adjustments of specific connectivities and time delays. If jitter builds up with the average number of synapses traversed and this is, in turn, roughly proportional to the time delay needed to encode a particular duration, then one has a potential explanation for the constant Weber fractions that are observed in discriminations of rhythms and other time intervals [52]. A theory of timing relations in arbitrary conduction networks would clearly be helpful.

This present treatment of timing networks barely ventures beyond an outline of the idea and what kinds of operations might potentially be carried out. Certainly, inhibitory inputs and anticoincidence operations need to be incorporated into such networks, and feedforward

and recurrent architectures need to be combined. Once these primitive networks are developed more fully and their behavior understood more deeply, then more realistic psychoneural models can be entertained that point to empirically testable hypotheses that address the real workings of the brain.

## 9. Conclusions

In the auditory nerve there is an abundance of temporal information that precisely and robustly encodes many perceptually relevant aspects of acoustic stimuli: periodicity, spectral shape, speech modulations, rhythms, and still longer time patterns. Most central models of auditory processing that utilize this timing information have assumed that a time-to-place transformation must occur in the ascending auditory pathway, such that central representations of auditory forms are based on excitation profiles of frequency- or periodicity-tuned units. In these models, auditory discrimination and recognition is performed by comparing stored excitation profiles with incoming ones.

However, if neural mechanisms exist by which timing information can be preserved and stored centrally, then purely temporal analyses of similarity and difference can be carried out by temporal correlation operations. We have outlined two basic processing architectures that could realize such operations. A simple, feedforward neural timing architecture has been presented that utilizes coincidence detectors and tapped delay lines to perform cross-correlation and/or convolution operations on two sets of inputs. Only those periodicities that are common to both inputs appear in the time structure of the outputs. The array functions as a temporal sieve whose summary autocorrelation function is the product of the autocorrelations of its inputs. To the extent that time structure of inputs reflect those of stimuli, such arrays can compute pitch similarity irrespective of timbre and timbral similarity independent of pitch. A simple recurrent timing architecture consisting of an array of many different delay loops is presented that amplifies and separates recurring time patterns.

These purely temporal modes of analysis are carried out on population-wide bases that obviate the need for precise point-to-point connectivities, explicit measurement of local features and/or internal clocks. Timing nets constitute a new and general neural network strategy for performing a host of basic auditory computations: extraction of common periodicities, detection of recurrent time patterns and separation of auditory objects. While the examples considered here are very rudimentary, they nevertheless afford glimpses of the kinds of perceptual computations that might be realized using temporal codes and timing nets.

## Acknowledgments

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

## References

- [1] Abeles, M. *Corticonics*, Cambridge, Cambridge University Press, 1990.
- [2] Bilsen, F.A. "Pronounced binaural pitch phenomenon," *J. Acoust. Soc. Am.*, 59: 467-468, 1976.
- [3] Braitenberg, V. "Functional interpretation of cerebellar histology," *Nature*, 190: 539-540, 1961.
- [4] Cariani, P. "As if time really mattered: temporal strategies for neural coding of sensory information," *Communication and Cognition - Artificial Intelligence (CC-AI)*, 12: 161-229, 1995.
- [5] Cariani, P. "Physiological correlates of periodicity pitch in the cochlear nucleus [abstract]," *Assoc. Res. Otolaryn. Abstr.*, 128, 1995.

- [6] Cariani, P. "Population-interval models for binaural periodicity pitches " *Soc. Neurosci. Abstr.*, 22: 649, 1996.
- [7] Cariani, P. "Temporal coding of sensory information," *Computational Neuroscience: Trends in Research, 1997*, J. M. Bower (ed.), New York, Plenum, pp. 591-598, 1997.
- [8] Cariani, P. and B. Delgutte. "Interspike interval distributions of auditory nerve fibers in response to concurrent vowels with same and different fundamental frequencies [abstract]," *Assoc. Res. Otolaryngology. Abs.*, 373, 1993.
- [9] Cariani, P., Delgutte, B. and Tramo, M. "Neural representation of pitch through autocorrelation". *Proc. Audio Eng. Soc.*, New York, 1997, Preprint #4583 (L-3), 1997.
- [10] Cariani, P.A. and Delgutte, B. "Neural correlates of the pitch of complex tones. I. Pitch and pitch salience," *J. Neurophysiol.*, 76: 1698-1716, 1996.
- [11] Cariani, P.A. and Delgutte, B. "Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase-invariance, pitch circularity, and the dominance region for pitch," *J. Neurophysiol.*, 76: 1717-1734, 1996.
- [12] Cherry, C. "Two ears – but one world," *Sensory Communication.*, W. A. Rosenblith (ed.), New York, MIT Press/John Wiley, pp. 99-117, 1961.
- [13] Chistovich, L.A. "Central auditory processing of peripheral vowel spectra," *J. Acoust. Soc. Am.*, 77: 789 - 805, 1985.
- [14] Chistovich, L.A. and Malinnikova, T.G.. "Processing and accumulation of spectrum shape information over the vowel duration," *Speech Communication*, 3: 361-370, 1984.
- [15] Colburn, S. and Durlach, N.I. "Models of binaural interaction," *Handbook of Perception.*, E. C. Carterette and M. P. Friedman (ed.), New York, Academic Press, pp. 467-518, 1978.
- [16] Cramer, E.M. and Huggins, W.H. "Creation of pitch through binaural interaction," *J. Acoust. Soc. Am.*, 30: 413-417, 1958.
- [17] Culling, J.F., Summerfield, Q. and Marshall, D.H. "Dichotic pitches as illusions of binaural masking release I: Huggins' pitch and the binaural edge pitch," *J. Acoust. Soc. Am.*, 103: 3509-3526, 1998.
- [18] de Boer, E. "On the "residue" and auditory pitch perception," *Handbook of Sensory Physiology.*, W. D. Keidel and W. D. Neff (ed.), Berlin, Springer, pp. 479-583, 1976.
- [19] de Cheveigné, A. "Cancellation model of pitch perception," *J. Acoust. Soc. Am.*, 103: 1261-1271, 1998.
- [20] de Ribaupierre, F. "Acoustical information processing in the auditory thalamus and cerebral cortex," *The Central Auditory System.*, G. Ehret and R. Romand (ed.), New York, Oxford University Press, pp. 317-397, 1997.
- [21] Goldstein, J.L. and Sruлович, p. "Auditory-nerve spike intervals as an adequate basis for aural frequency measurement," *Psychophysics and Physiology of Hearing.*, E. F. Evans and J. P. Wilson (ed.), London, Academic Press, pp. 337-346, 1977.
- [22] Greenberg, S. *Neural Temporal Coding of Pitch and Vowel Quality: Human Frequency-Following Response Studies of Complex Signals*, Los Angeles, UCLA Working Papers in Phonetics #52, 1980.
- [23] Grossberg, S. "Neural dynamics of motion perception, recognition learning, and spatial attention," *Mind as Motion: Explorations in the Dynamics of Cognition.*, R. F. Port and T. van Gelder (ed.), Cambridge, MIT Press, pp. 449-490, 1995.
- [24] Hall III, J.W. and Peters, R.W. "Pitch for nonsimultaneous successive harmonics in quiet and noise," *J. Acoust. Soc. Am.*, 69: 509-513, 1981.
- [25] Hirahara, T., Cariani, P. and Delgutte, B. "Representation of low-frequency vowel formants in the auditory nerve," *Proc. ESCA Research Tutorial on The Auditory Basis of Speech Perception*, W.A. and S. Greenberg (eds), Keele University, U. K., pp. 83-86, 1996.
- [26] Houtsma, A.J.M. and Goldstein, J.L. "The central origin of the pitch of complex tones: Evidence from musical interval recognition," *J. Acoust. Soc. Am.*, 51: 520-529, 1972.
- [27] Jeffress, L.A. "A place theory of sound localization," *J. Comp. Physiol. Psychol.*, 41: 35-39, 1948.
- [28] John, E.R. "Switchboard vs. statistical theories of learning and memory," *Science*, 177: 850-864, 1972.
- [29] John, E.R. and Schwartz, E.L. "The neurophysiology of information processing and cognition," *Ann. Rev. Psychol.*, 29: 1-29, 1978.
- [30] Jones, M.R. "Time, our lost dimension: toward a new theory of perception, attention, and memory," *Psychological Review*, 83: 323-255, 1976.
- [31] Jones, M.R. and Yee, W. "Attending to auditory events: the role of temporal organization," *Thinking in Sound: The Cognitive Psychology of Human Audition.*, S. McAdams and E. Bigand (ed.), Oxford, Clarendon Press, pp. 69-112, 1993.

- [32] Langner, G. "Periodicity coding in the auditory system," *Hear. Res.*, 60: 115-142, 1992.
- [33] Licklider, J.C.R. "A duplex theory of pitch perception," *Experientia*, VII: 128-134, 1951.
- [34] Licklider, J.C.R. "Three auditory theories," *Psychology: A Study of a Science*, S. Koch (ed.), New York, McGraw-Hill, pp. 41-144, 1959.
- [35] Longuet-Higgins, H.C. *Mental Processes: Studies in Cognitive Science.*, Cambridge, MIT Press, 1987.
- [36] Longuet-Higgins, H.C. "A mechanism for the storage of temporal correlations," *The Computing Neuron.*, R. Durbin, C. Miall and G. Mitchison (ed.), Wokingham, Addison-Wesley, pp. 99-104, 1989.
- [37] Lyon, R. and Shamma, S. "Auditory representations of timbre and pitch," *Auditory Computation.*, H. Hawkins, T. McMullin, A. N. Popper and R. R. Fay (eds.), New York, Springer Verlag, pp. 517, 1995.
- [38] MacKay, D.M. "Self-organization in the time domain," *Self-Organizing Systems 1962.*, M. C. Yovitts, G. T. Jacobi and G. D. Goldstein (eds.), Washington, D.C., Spartan Books, pp. 37-48, 1962.
- [39] Meddis, R. and Hewitt, M.J. "Virtual pitch and phase sensitivity of a computer model of the auditory periphery.I.Pitch identification," *J. Acoust. Soc. Am.*, 89: 2866-2882, 1991.
- [40] Meddis, R. and Hewitt, M.J. "Modeling the perception of concurrent vowels with different fundamental frequencies," *J. Acoust. Soc. Am.*, 91: 233-245, 1992.
- [41] Meyer-Eppler, W. "Exhaustion methods of selecting signals from noisy backgrounds," *Communication Theory.*, W. Jackson (ed.), London, Butterworths, pp.183-194, 1953.
- [42] Miller, R.R. and Barnet, R.C. "The role of time in elementary associations," *Current Directions in Psychological Science*, 2: 106-111, 1993.
- [43] Moore, B.C.J. *Introduction to the Psychology of Hearing*, London, Academic, 1997.
- [44] Morrell, F. "Electrical signs of sensory coding," in *The Neurosciences: A Study Program.*, G. C. Quarton, T. Melnechuck and F. O. Schmitt (ed.), New York, Rockefeller U. Press, pp. 452-469, 1967.
- [45] Palmer, A.R. "Segregation of the responses to paired vowels in the auditory nerve of the guinea pig using autocorrelation," in *The Auditory Processing of Speech.*, M. E. H. Schouten (ed.), Berlin, Mouton de Gruyter, pp. 115-124, 1992.
- [46] Patterson, R.D., Allerhand, M.H. and Giguere, C. "Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform," *J. Acoust. Soc. Am.*, 98: 1890-1894, 1995.
- [47] Perckell, D.H. and Bullock, T.H. "Neural Coding," *Neurosci. Res. Prog. Bull.*, 6: 221-348, 1968.
- [48] Rees, A. and Møller. A.R. "Stimulus properties influencing the responses of inferior colliculus neurons to amplitude-modulated sounds," *Hearing Res.*, 27: 129-143, 1987.
- [49] Rees, A. and Palmer. A.R. "Neuronal responses to amplitude-modulated and pure-tone stimuli in the guinea pig inferior colliculus, and their modification by broadband noise," *J. Acoust. Soc. Am.*, 85: 1978-1994, 1989.
- [50] Reitboeck, H.J. "Neural mechanisms of pattern recognition," in *Sensory Processing in the Mammalian Brain.*, J. S. Lund (ed.), Oxford, Oxford U. Press, pp. 307-330, 1989.
- [51] Rhode, W.S. "Interspike intervals as correlates of periodicity pitch in cat cochlear nucleus," *J. Acoust. Soc. Am.*, 97: 2414-2429, 1995.
- [52] Rosenbaum, D.A. "Broadcast theory of timing," in *Timing of Behavior: Neural, Psychological and Computational Perspectives.*, D. A. Rosenbaum and C. E. Collyer (ed.), Cambridge, MA, MIT Press, pp. 215-235, 1998.
- [53] Schreiner, C.E. and Langner, G. "Coding of temporal patterns in the central auditory nervous system," in *Auditory Function: Neurobiological Bases of Hearing.*, (ed.), New York, Wiley, pp. 337-361, 1988.
- [54] Schwartz, D.W.F. and Tomlinson, R.W.W. "Spectral response patterns of auditory cortex neurons to harmonic complex tones in alert monkey (*Macaca mulatta*)," *J. Neurophys.*, 64: 282-298, 1990.
- [55] Siebert, W.M. "Frequency discrimination in the auditory system: place or periodicity mechanisms?" *Proc IEEE*, 58: 723-730, 1970.
- [56] Slaney, M. and Lyon, R.F. "On the importance of time - a temporal representation of sound," in *Visual Representations of Speech Signals.*, M. Cooke, S. Beet and M. Crawford (ed.), New York, John Wiley, pp. 95-118, 1993.
- [57] Steinschneider, M., Reser, D.H., Fishman, Y.I. and Arezzo, J. "Click train encoding in primary auditory cortex of the awake monkey: Evidence for two mechanisms subserving pitch perception," *J. Acoust. Soc. Am.*, 104: 2395-2955, 1998.
- [58] Thatcher, R.W. and John, E.R. *Functional Neuroscience, Vol. I. Foundations of Cognitive Processes.*, Hillsdale, NJ, Lawrence Erlbaum, 1977.
- [59] Yellott, J.I. and Iverson, G.J. "Uniqueness properties of higher-order autocorrelation functions," *J. Optic. Soc. Am. A*, 9: 388-404, 1992.