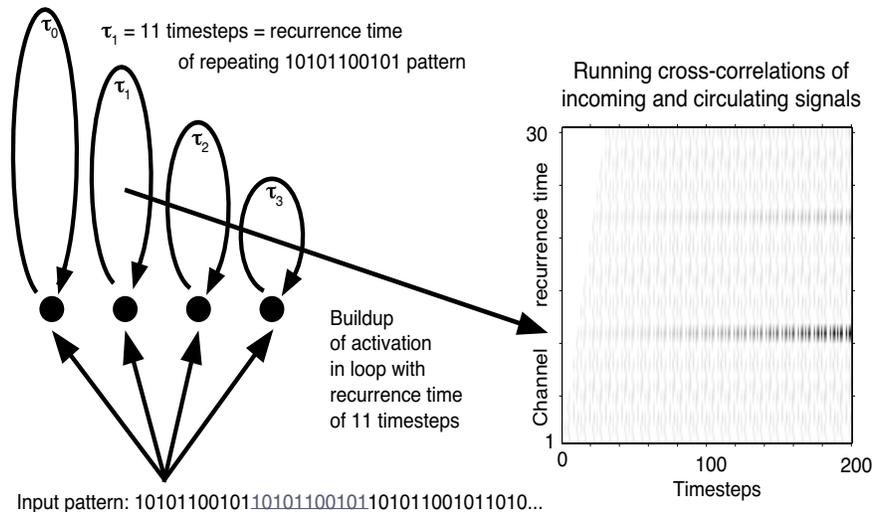


Papers on Neural Timing Nets (2001-2016)

Peter Cariani



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

Cariani, P. A. (2001). Neural timing nets. *Neural Networks*, 14(6-7), 737-753.

Cariani, P. (2002). Temporal codes, timing nets, and music perception. *J. New Music Res.*, 30(2), 107-136.

Cariani, P., & Leman, M. (2002). Introduction: Music and time. *J. New Music Res.*, 30(2), 105.

Cariani, P. (2004). Temporal codes and computations for sensory representation and scene analysis. *IEEE Transactions on Neural Networks, Special Issue on Temporal Coding for Neural Information Processing*, 15(5), 1100-1111.

Cariani, P. (2005). Recurrent timing nets for F0-based speaker separation. In P. Divenyi (Ed.), *Perspectives on Speech Separation* (pp. 31-53). New York: Kluwer Academic Publishers.

Cariani, P. (2012). Creating new primitives in minds and machines. In J. McCormack & M. D'Inverno (Eds.), *Computers and Creativity* (pp. 395-430). New York: Springer.

Cariani, P. (2015). Outline of a cybernetic theory of brain function based on neural timing nets. *Kybernetes*, 44(8/9), 1219-1232.

Cariani, P. (2016, December 8-10). *Time is of the essence*. Paper presented at the A Body of Knowledge – Embodied Cognition and the Arts, University of California at Irvine.

Cariani, P. (2017). Temporal memory traces as anticipatory mechanisms. In M. Nadin (Ed.), *Anticipation and Medicine* (pp. 105-136). Switzerland: Springer.

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 timbre of stationary sounds, such as vowel quality [4][9][25][37][40][45]. These timbres are 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, how does it use 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 about 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 AM 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].

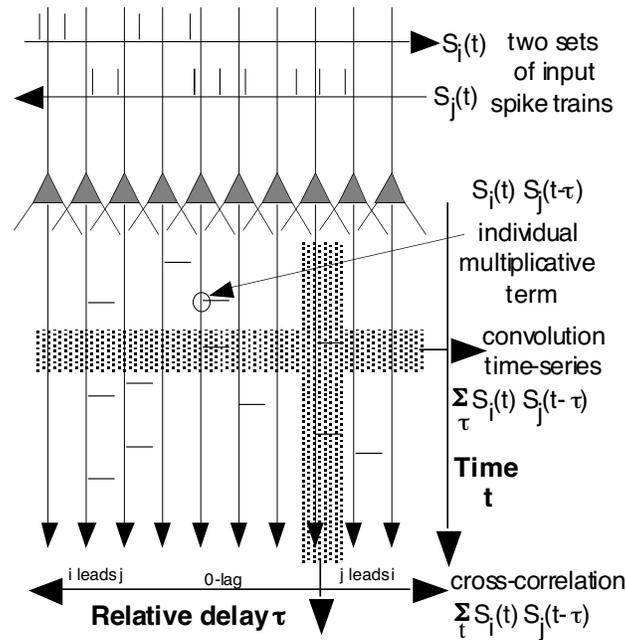


Figure 1 Simple feedforward timing net consisting of an array of coincidence detectors and two sets of tapped delay lines through which input signals S_i and S_j arrive.

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, $S_i(t) * S_j(t - \tau)$). 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, τ_0 , and the other has N such intervals, then $M * N \tau_0$ intervals will appear in the outputs. Within the temporal

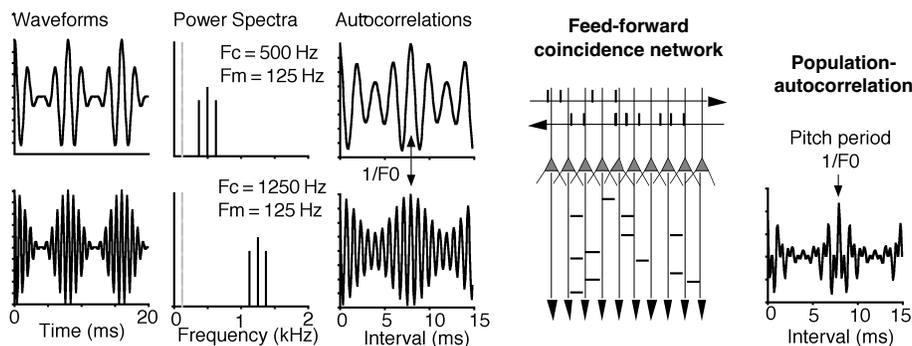


Figure 2 Effect of passing two signals through the coincidence array. The stimuli are two AM tones with different carriers ($f_c = 500$ Hz, 1250 Hz) but the same modulation frequency ($f_m = 125$ Hz). The AM tones have no harmonics in common, but they produce a common low pitch at their “missing fundamental”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

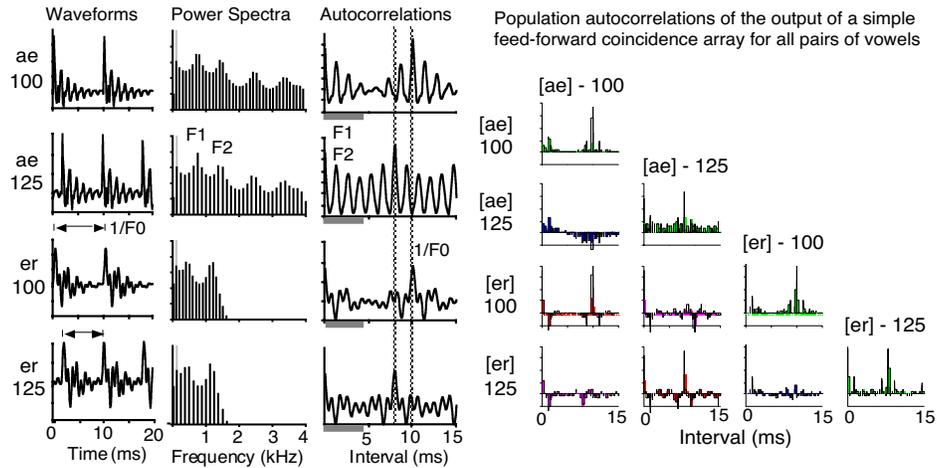


Figure 3 Left: Waveforms, power spectra, and autocorrelation functions for four vowels. The vowel set consists of combinations of two different fundamental frequencies ($f_0 = 100, 125$ Hz) and two formant structures. Horizontal arrows above waveforms and vertical lines in autocorrelations indicate fundamental 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 [e^r] (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 [e^r] (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

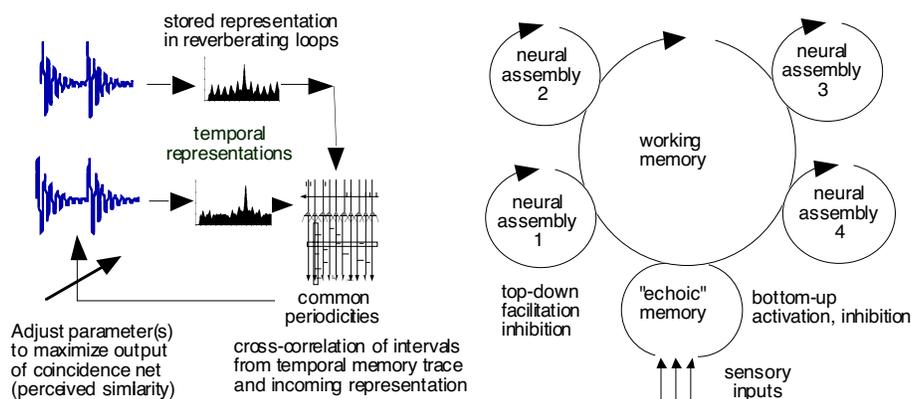


Figure 4 Temporal memory traces, matching tasks, and the build-up of perceptual forms

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]),

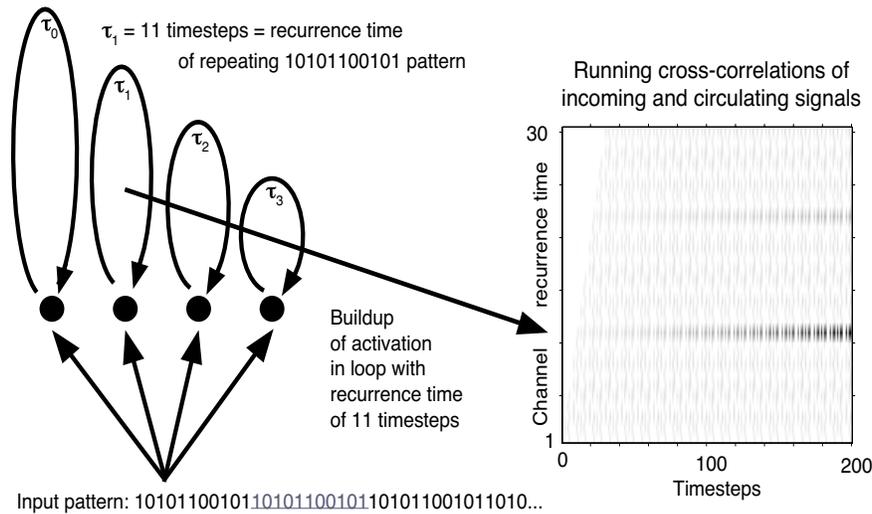


Figure 5 Behavior of a simple recurrent timing net for periodic pulse-train patterns.

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.

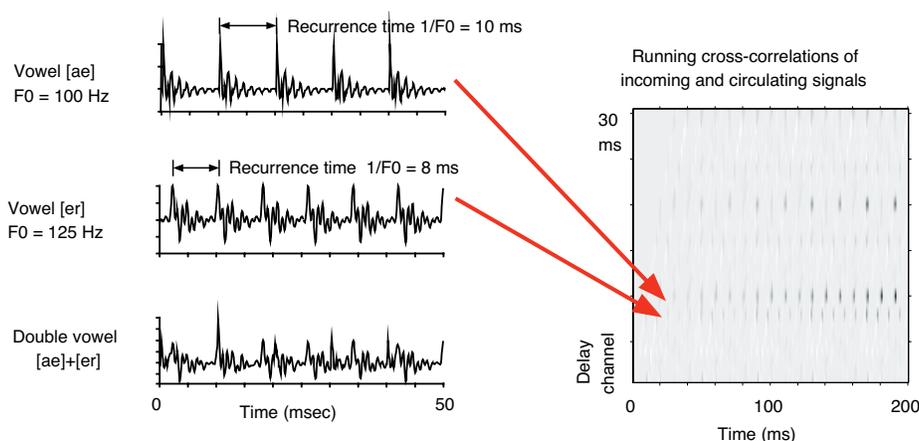


Figure 6 Separation of two auditory objects, with differential fundamental frequencies, in a simple recurrent net.

Both feed-forward 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." In *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." In *Computational Neuroscience: Trends in Research, 1997*, J. M. Bower (ed.), New York: Plenum, pp. 591–598, 1997.
- [8] Cariani, P. and Delgutte, B. "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.* (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." In *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." In *Handbook of Perception.*, E. C. Carterette and M. P. Friedman (eds.), 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." In *Handbook of Sensory Physiology.*, W. D. Keidel and W. D. Neff (eds.), Berlin: Springer-Verlag, 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." In *The Central Auditory System.*, G. Ehret and R. Romand (eds.), 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." In *Psychophysics and Physiology of Hearing.*, E. F. Evans and J. P. Wilson (eds.), 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*. UCLA Working Papers in Phonetics #52, 1980.
- [23] Grossberg, S. "Neural dynamics of motion perception, recognition learning, and spatial attention." In *Mind as Motion: Explorations in the Dynamics of Cognition.*, R. F. Port and T. van Gelder (eds.), 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*, S. Greenberg and W. A. Ainsworth (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." *Psych. Rev.*, 83: 323–255, 1976.
- [31] Jones, M. R. and Yee, W. "Attending to auditory events: The role of temporal organization." In *Thinking in Sound: The Cognitive Psychology of Human Audition.*, S. McAdams and E. Bigand (eds.), 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*, 7: 128–134, 1951.
- [34] Licklider, J. C. R. "Three auditory theories." In *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, MA: MIT Press, 1987.
- [36] Longuet-Higgins, H. C. "A mechanism for the storage of temporal correlations." In *The Computing Neuron*, R. Durbin, C. Miall and G. Mitchison (eds.), Wokingham, MA: Addison-Wesley, pp. 99–104, 1989.
- [37] Lyon, R. and Shamma, S. "Auditory representations of timbre and pitch." In *Auditory Computation*, H. Hawkins, T. McMullen, A. N. Popper and R. R. Fay (eds.), New York: Springer Verlag, pp. 221–270, 1995.
- [38] MacKay, D. M. "Self-organization in the time domain." In *Self-Organizing Systems*, M. C. Yovitts, G. T. Jacobi and G. D. Goldstein (eds.), Washington, DC: 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." In *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. *An Introduction to the Psychology of Hearing* (3rd ed.). 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 (eds.), New York: Rockefeller University 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] Perkell, 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 University 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 (eds.), 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*, G. Edelman, W. E. Gall, and W. M. Cowan (eds.), 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 (eds.), New York: 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, Volume 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.



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. © 2001 Elsevier Science Ltd. All rights reserved.

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; Perrell & 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).

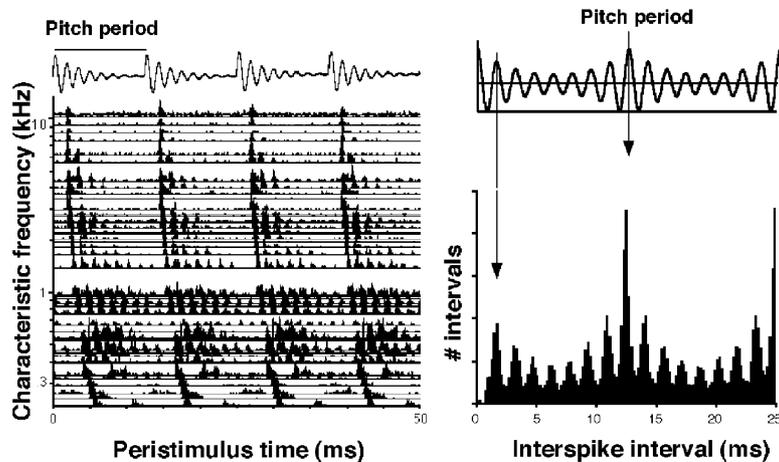


Fig. 1. Temporal coding of pitch and timbre in the auditory nerve. Top: Stimulus waveform, Single formant vowel, $F_0 = 80$ Hz, $F_1 = 640$ Hz, 60 dB SPL, 100 presentations/fiber. Left, peristimulus time histograms of the responses of 52 auditory nerve fibers of Dial-anesthetized cats, arranged by fiber 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 ($1/F_0 = 1/80$ 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 electroreception ($<1 \mu\text{s}$) microsecond; bat echolocation, ($<1 \mu\text{s}$ 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 infor-

mation 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 & Sruлович, 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 subservise. 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 . $CCF(D_k) = \sum S_i(t) * S_j(t - D_k)$, 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). $Conv(t) = \sum S_i(t) * S_j(t - D_k)$, 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 $AC_k(\tau) = \sum S_i(t) * S_j(t - D_k) * \sum S_i(-\tau) * S_j(t - \tau - D_k)$ 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, $SAC(\tau) = \sum AC_k(\tau)$. The population-interval distribution

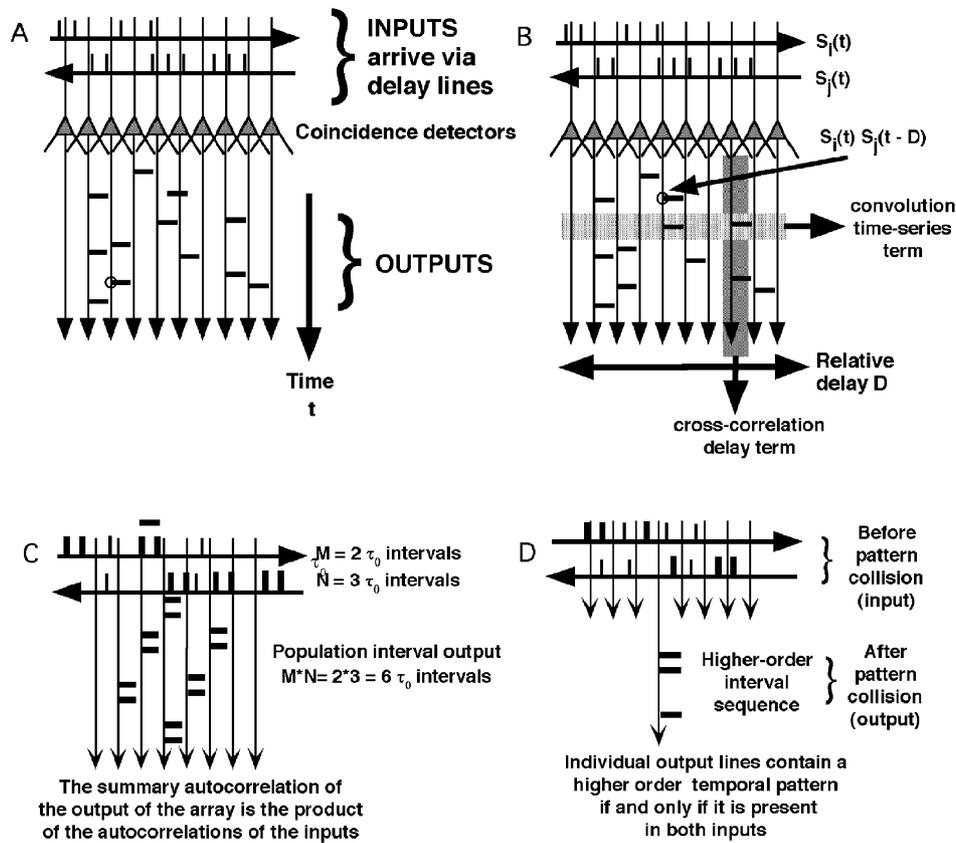


Fig. 2. A simple feedforward timing net. (a) General schematic of a coincidence array traversed by tapped delay lines. Each coincidence element (pyramidal cell icons) receives two inputs. The spatial position of a coincidence element in the array determines the relative time of arrival of the two inputs (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 ($F_0 = F_m = 125$ Hz), but different carrier frequencies ($F_c = 500$ 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, $F_0 = 125$ 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 μ s 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 ($1/F_0 = 8$ 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. Effec-

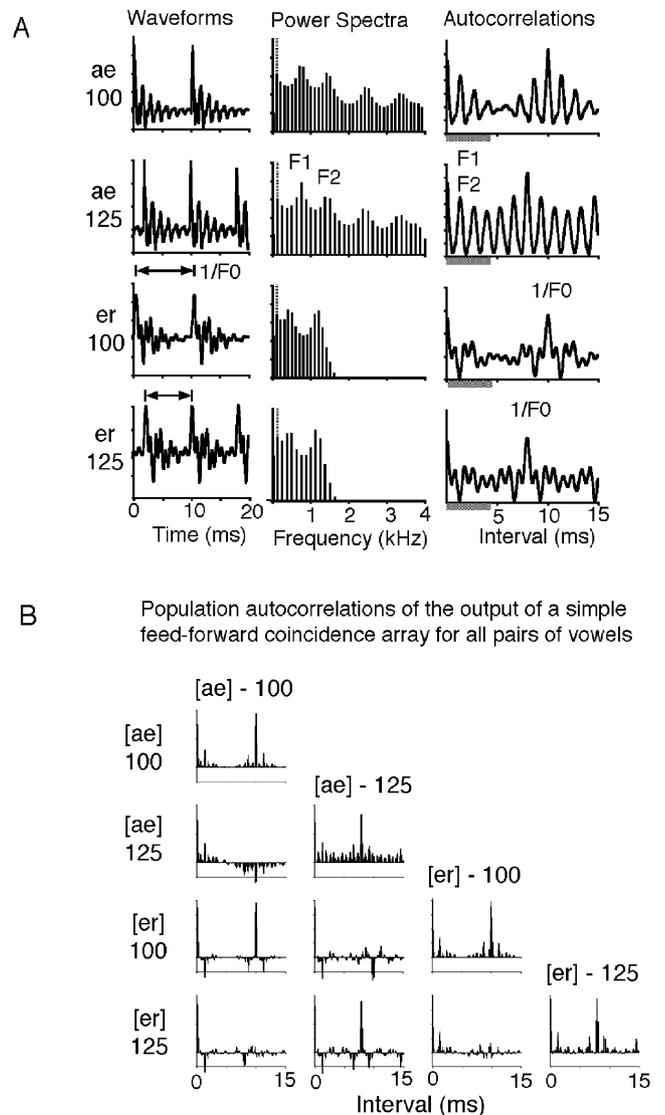


Fig. 3. (a) Waveforms, power spectra, and autocorrelation functions for four vowels. The vowel set consists of combinations of two different fundamental frequencies ($F_0 = 100, 125$ Hz) and two formant structures. Horizontal arrows above waveforms and vertical lines in autocorrelations indicate fundamental periods ($1/F_0 = 8, 10$ ms), which correspond to voice pitch periods. Shaded bars indicate periodicities associated with formant structure that give rise to differences in vowel quality (timbre). (b) Population autocorrelations of the output of the coincidence array for all vowel pairs.

tively, the network extracts the common fundamental periodicity without ever making any sort of explicit estimation of the fundamentals of the two input signals.

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

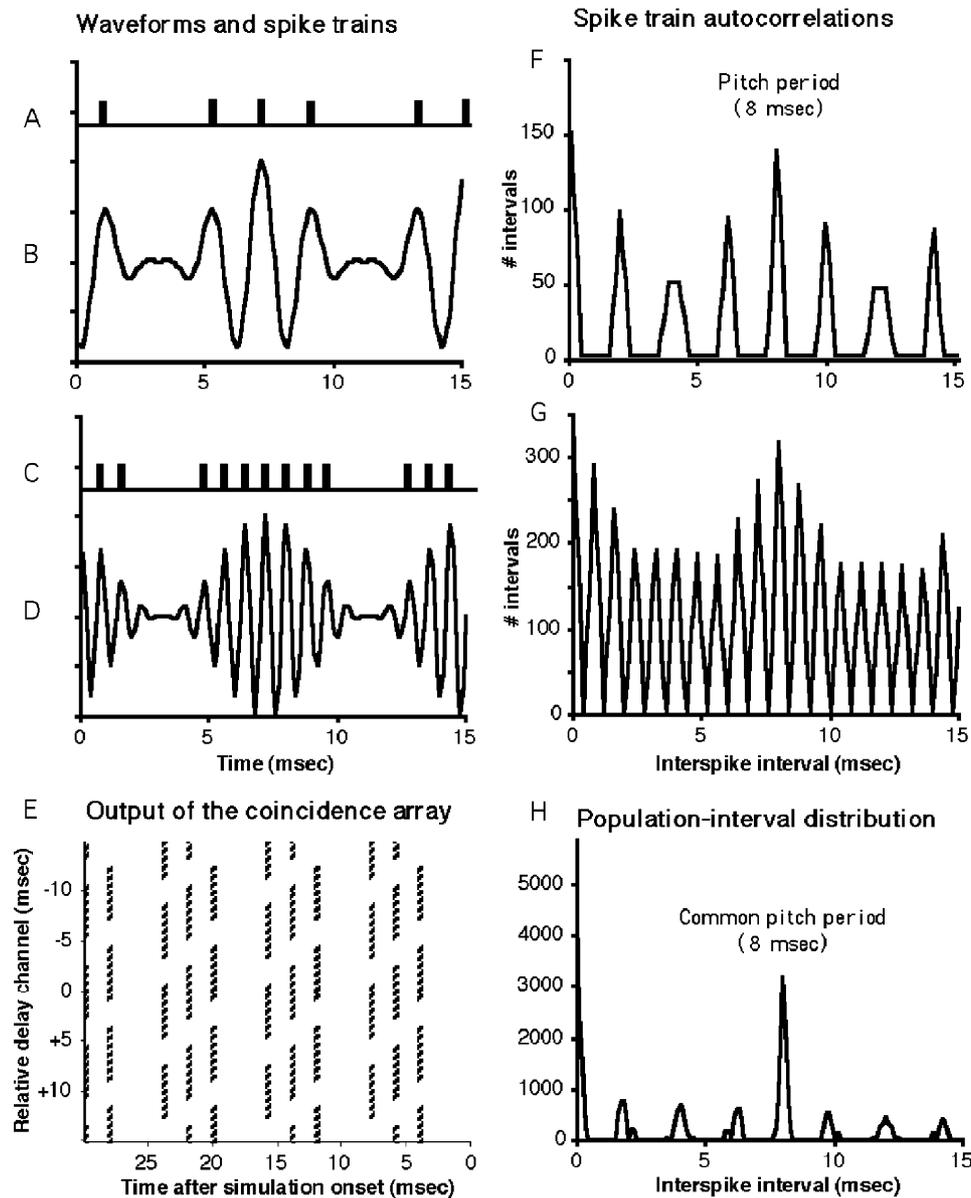


Fig. 4. Extraction of a common fundamental frequency by a coincidence array. (a) and (c) Pulse trains derived from maxima of two waveforms. Pulses are $300 \mu\text{s}$ wide. (b) and (d) Waveforms of two amplitude-modulated (AM) tones with different carriers ($F_c = 500, 1250$ Hz) but same modulation frequency ($F_m = 125$ 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' ($F_0 = F_m = 125\text{Hz}$). (e) Output of a coincidence array in response to presentation of the two pulse trains. The effective coincidence window was $600 \mu\text{s}$. The output of the coincidence array is shown as a function of the relative delay channel (ordinate) and time (abscissa). (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

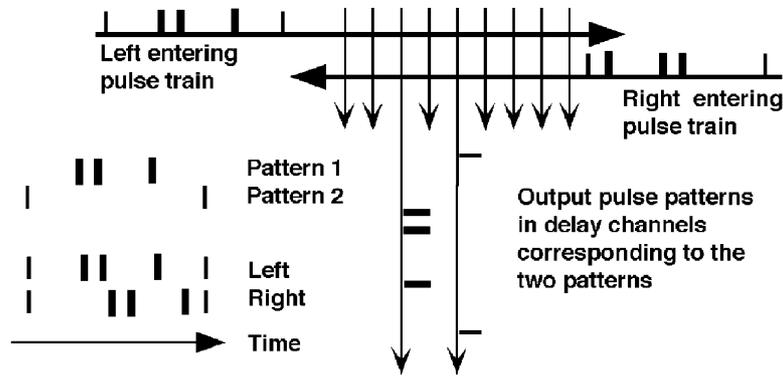


Fig. 5. Fusion and separation of patterns by relative arrival time disparities in a feedforward timing net. Two temporal patterns of pulses are shown that have different inter-channel arrival times, such that the temporal disparity between the two patterns is different in each channel. As a consequence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

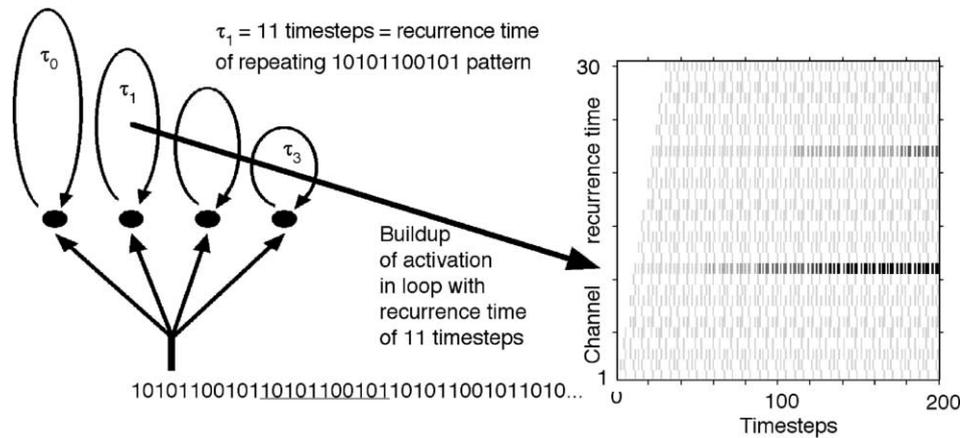


Fig. 6. Behavior of a simple recurrent timing net for periodic pulse train patterns. Left, recurrent timing net consisting of an array of coincidence detectors with associated loops having a range of recurrence times. Right, output of the coincidence array, arranged by the length of delay loop (ordinate)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

In addition, 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

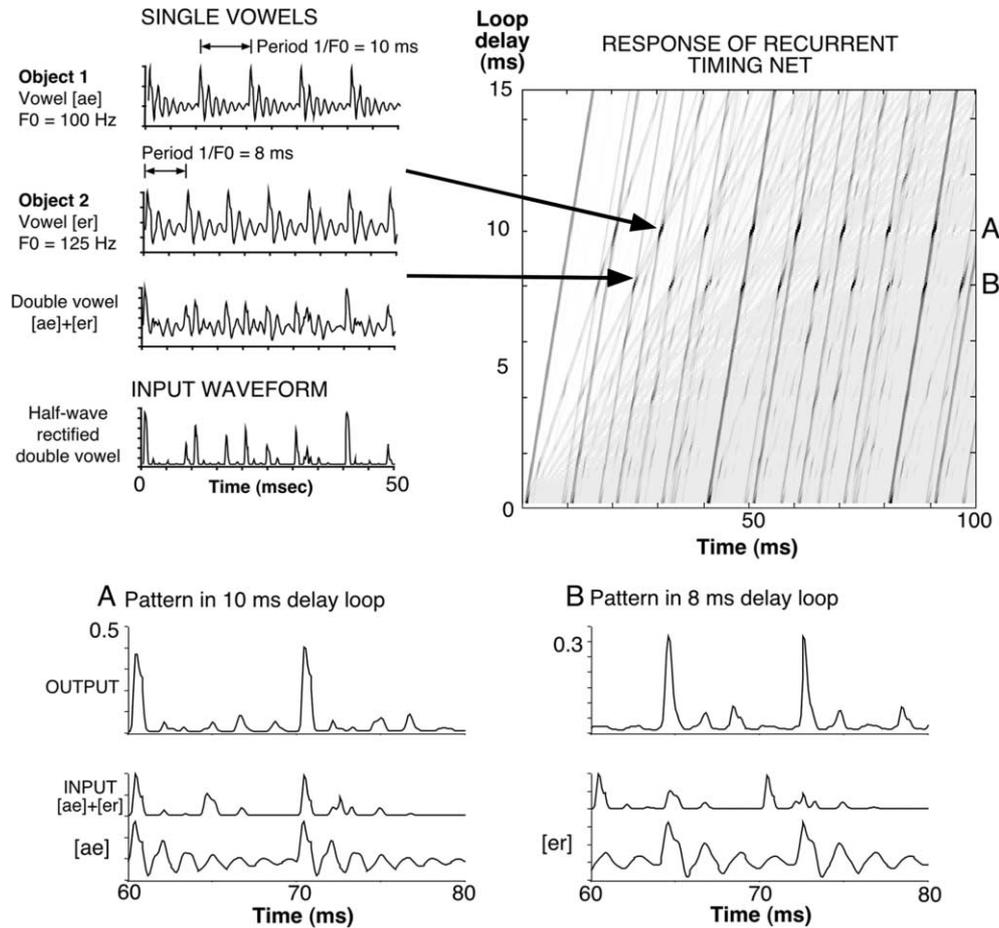


Fig. 7. Separation of auditory objects through temporal pattern coherence. Top waveform plots. Two synthetic vowels [ae] and [er] with different fundamental frequencies 100 and 125 Hz were summed together to form a double vowel. The composite waveform was half-wave rectified and presented to a recurrent 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 (abscissa), 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 repeat-

ing 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 μ s 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

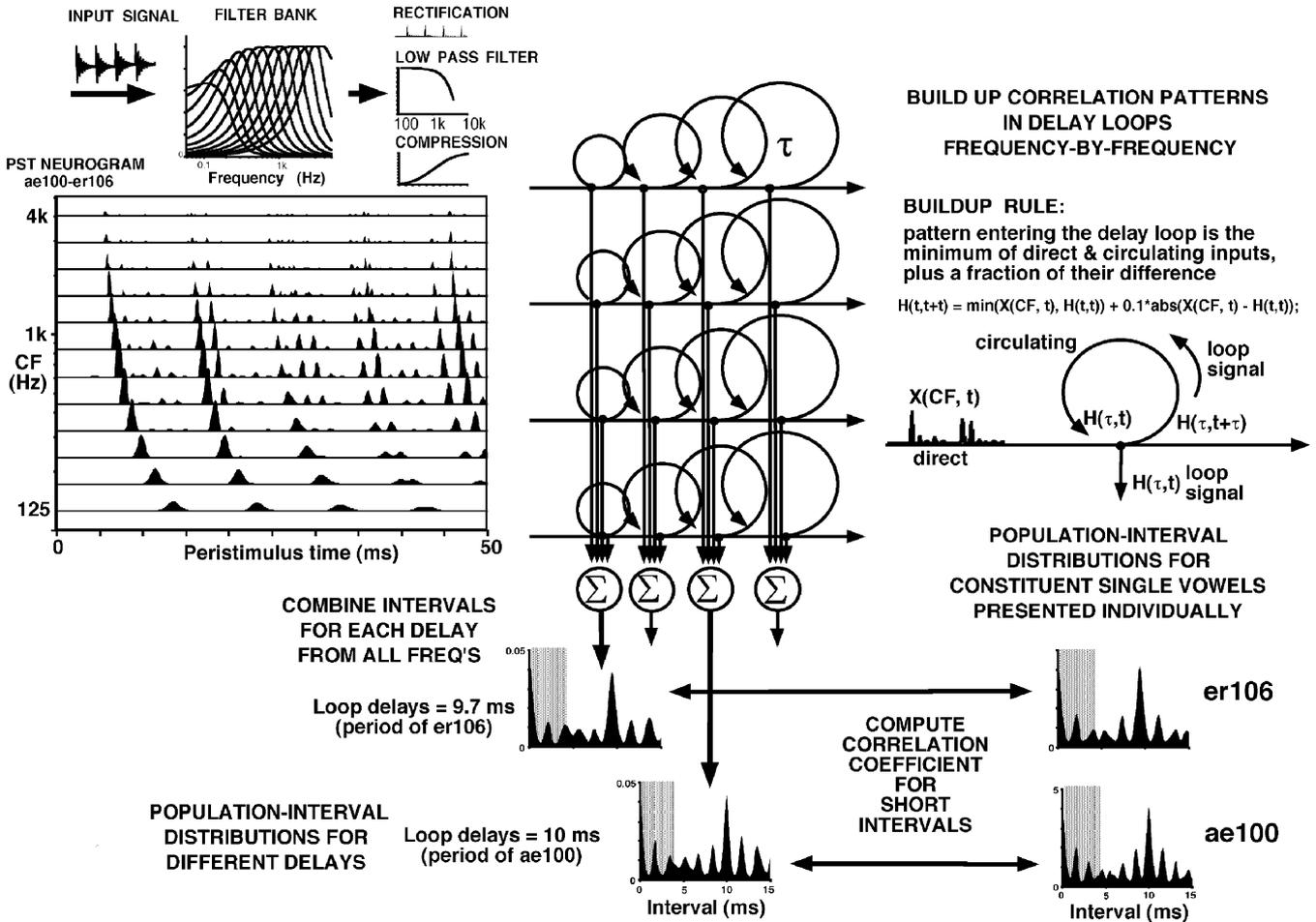


Fig. 8. Multichannel recurrent timing net. Schematic for multichannel, frequency-based recurrent timing net. Top left plots show stages in the auditory nerve simulation: band-pass filtering, half-wave rectification, low pass filtering, and rate compression. The simulation generates a PST neurogram (response 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,

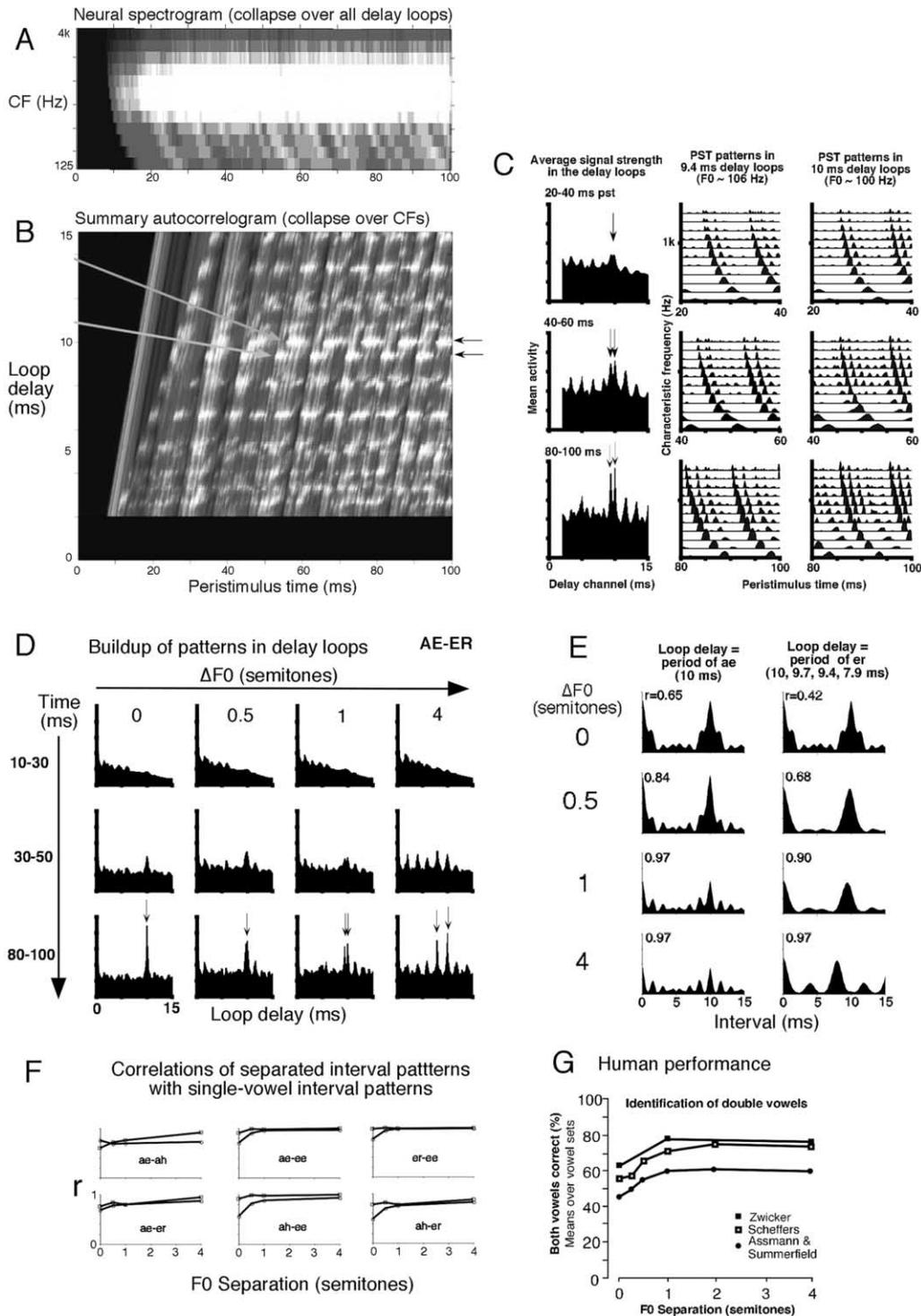


Fig. 9. Separation of concurrent double vowels using a multichannel recurrent timing net. Response to the double vowel ae–er, $F_0(\text{ae}) = 100$ Hz, $F_0(\text{er}) = 106$ Hz. (a) Plot of mean activation vs. time as a function of characteristic frequency (combining across loop delays). (b) Mean activation 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 F_0 separation. (g) Percent correct of double vowels whose constituents were correctly identified by human listeners as a function of F_0 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 $F0$ -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. $F0$ -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 iden-

tity, 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 an 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 systematic Sensory, 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 small 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. 1). 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 Bekeesy, 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.

Introduction: Music and Time

Music is nothing if not the designed creation of temporal patterns. As objects of perception and cognition, musical forms range widely in complexity, from the simplest rhythms, melodies, and harmonies, to some of the most elaborate patterns that can be comprehended by human beings. As an art, the ultimate ends of music are related to expression and evocation of emotion. Both the perceptual-cognitive and aesthetic-hedonic aspects of music must have their correlates in neural informational processes embedded in biological brains.

Musical form exists both in time and of time. Successions of sonic events unfold over time to create temporal patterns and expectancies. Music plays on the perceptual and cognitive salience of temporal patterns. Anticipations are formed and violated through repetition and change to create cycles of tension and resolution.

Music also exists of time. Basic primitive qualities of pitch, timbre, and rhythm may themselves be represented in the brain as temporal patterns. The neural mechanisms for apprehending harmonic relations in tonal music may therefore be the consequences of the basic neural codes and computations that the brain uses to analyze the auditory environment. Thus not only can there be representations of sequences and patterns of events in time, but also that musical objects themselves may be fundamentally temporal in nature.

This issue explores modeling of music and time from different points of view. Peter Cariani, in “Temporal Codes, Timing Nets, and Music Perception”, introduces the notion of timing nets as neural structures for temporal information processing on neural spike trains. Feedforward timing nets are shown to be able to extract common periodicities underlying pitch and timbre, while recurrent nets are used to build up and extract multiple rhythmic subpatterns.

In his “Real-time Recognition of Improvisations with Adaptive Oscillators and a Recursive Bayesian Classifier” Petri Toiviainen gives an overview of his work in tempo tracking and beat induction, which is then applied to a real-time MIDI interactive music system for the recognition and accompaniment of tonal improvisations.

William Sethares and Thomas Staley, in “Meter and Periodicity in Musical Performance”, present the periodicity transform as a mathematical tool to locate periodicities in musical audio signals that projects temporal event patterns onto a set of (non-orthogonal) periodic subspaces. The method detects periodicities that correspond to the pulse, the measure, and phrases.

Masataka Goto, in “An Audio-Based Real-Time Beat Tracking System for Music with or without Drum-sounds and its Applications”, describes a real-time beat tracking system that recognizes a hierarchical beat structure of quarter-note, half-note, and measure level in real-world audio signals. The model uses frequency analysis based on top-down information, as well as knowledge concerning the presence or absence of drum-sounds.

Edward Large, in “Periodicity, Pattern Formation, and Rhythmic Structure”, describes an approach to metrical structure using a self-organized, dynamic structure composed of self-sustaining oscillators. The structure anticipates futures events and attempts to capture the perception of metrical structure through formation of stable patterns of interaction that implement dynamic attentional strategies.

Douglas Eck, in “A Positive-Evidence Model for Classifying Rhythmical Patterns”, presents a model that predicts the beat and pattern complexity in rhythmical patterns. Several datasets are used to examine model predictions and to compare them with existing models.

In this issue we have attempted to bring together these different aspects of music and time. The idea for this special issue on musical periodicity and rhythm perception came as a result of a workshop on Music and Timing Networks held on 29 October 1999 at the Institute for Psychoacoustics and Electronic Music (IPEM) of Ghent University. The editors wish to thank the Research Society for the Foundations of Music Research of the Belgian Fund for Scientific Research for their generous support of this workshop, which made this issue possible.

Peter Cariani and Marc Leman

Temporal Codes, Timing Nets, and Music Perception

Peter Cariani

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

Abstract

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

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

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

Introduction

Music entails the temporal patterning of sound for pleasure. As such, it involves the generation of simple and complex temporal patterns and expectancies over many different time scales. Music engages both the texture of auditory qualities and the general time sense. On the shortest, millisecond time scales, periodic acoustic patterns evoke qualities of pitch and timbre, while longer patterns create rhythms and larger musical structures. How neural mechanisms in the brain subservise these perceptual qualities and cognitive structures are

questions whose answers are not currently known with any degree of clarity or precision.

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

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

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

Accepted: 10 July, 2001

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Temporal coding of auditory forms

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

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

Population-interval distributions as auditory representations

An account of the neural coding of the pitch of individual musical notes is fundamental to understanding their concurrent and sequential interactions, the vertical and hori-

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

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

The two aspects of neural response, cochlear place and time, can be seen in Figure 1. The acoustic stimulus is a synthetic vowel whose fundamental frequency (F_0) is 80 Hz. Its waveform, power spectrum, and autocorrelation function are respectively shown in panels A, C, and D. Spike trains of single auditory nerve fibers of anesthetized cats were recorded in response to 100 presentations of the stimulus at a moderate sound pressure level (60 dB SPL) (Cariani & Delgutte, 1996a). The “neurogram” (B) shows the post-stimulus time (PST) histograms of roughly 50 auditory nerve fibers. These histograms plot the probability of occurrence of spikes at different times after the stimulus onset. The most striking feature of the neurogram is the widespread nature of the temporal discharge patterns that are associated with the periodicity of the fundamental. Even fibers whose characteristic frequencies are well above the formant frequency of 640 Hz, around which virtually all of the spectral energy of the stimulus lies, nevertheless convey pitch information. The widespread character of temporal patterns across cochlear frequency territories is a consequence of the broad nature of the low-frequency tails of tuning curves (Kiang et al., 1965). The profile of average driven discharge rates are shown in panel D. The driven rate is the firing rate of a fiber under acoustical stimulation minus its spontaneous discharge rate in quiet. In order to initiate a spectral pattern analysis for estimating the pitch of this vowel, a rate-place representation would have to resolve the individual partials of the stimulus (the harmonics in panel C, which are plotted on the same log-frequency scale as D). In practice, discharge rates of cat auditory nerve fibers provide very poor resolution of the individual harmonics of complex tones, even at very low harmonic numbers. Thus, while there is a coarse tonotopic

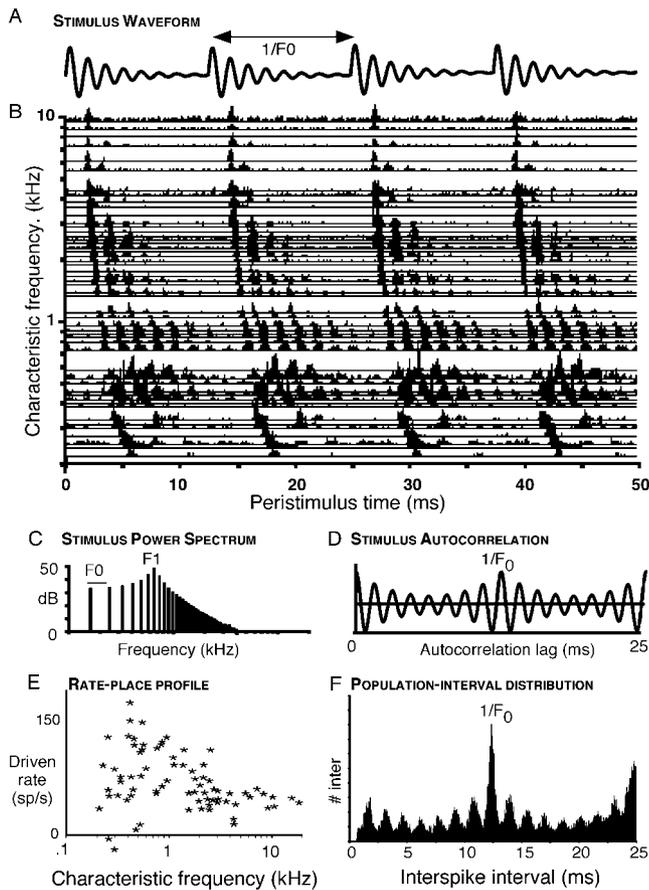


Fig. 1. Temporal coding of musical pitch in the auditory nerve. Auditory nerve responses to a harmonic complex tone with a single formant. (a) Stimulus waveform. A strong, low voice pitch is heard at the fundamental ($F_0 = 80$ Hz, pitch period (double arrow) $1/F_0 = 12.5$ ms). (b) Peristimulus time histograms of cat auditory nerve fibers (100 presentations at 60 dB SPL). Histogram baselines indicate fiber characteristic frequencies (CF's). (c) Stimulus power spectrum. (d) Stimulus autocorrelation function. (e) Rate-place profile, driven rates as a function of CF. (f) Population-interval distribution formed by summing all-order intervals from all fibers. For further details, see Cariani (1999).

pattern of activity present if one orders the fibers by their characteristic frequencies (cochlear place), this organization is not precise enough to subserve the pitch of complex tones. In contrast, interspike interval information from even a handful of auditory nerve fibers is sufficient to yield reasonably accurate estimates of the fundamental. Pooling interval information from many fibers integrates information from all frequency regions and yields still more precise representations. The population-interval distribution (F) of the ensemble of fibers is formed by pooling all of the interspike intervals from the spike trains produced by the individual fibers. These interspike intervals include time intervals between successive and nonsuccessive spikes, i.e., both "first-order" and "higher-order" intervals are pooled together to form "all-order" interval distributions. Making histograms of all-order intervals is formally equivalent to computing the

autocorrelation of a spike train. The population interval histogram (F) shows a very clear peak that corresponds to the fundamental period. For harmonic complexes such as this, the voice pitch that is heard would be matched to a pure tone with the same period, i.e., the pitch is heard at the fundamental frequency. Because of cochlear filtering and phase-locking, the form of the population-interval distribution (F) resembles that of the stimulus autocorrelation function (D) (Cariani, 1999). On the basis of such histograms that contain on the order of 5000 intervals, the fundamental period for such a harmonic tone complex can be reliably estimated, with a standard error of less than 1% (Cariani & Delgutte, 1996a).

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

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

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

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

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

Simulated population-interval distributions

We are interested in how population-interval representations associated with different music notes might be related to each other. A computer simulation of an array of auditory nerve fibers was used to make systematic comparisons between the population-interval distributions that would be produced by different musical sounds. The purpose of the simulation is to replicate the essential temporal features of the auditory nerve response to steady-state signals at moderate to high sound pressure levels in a computationally efficient manner. The MATLAB simulation incorporated bandpass filtering, half-wave rectification, low pass filtering, and rate compression

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

Population-interval distributions and autocorrelation

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

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

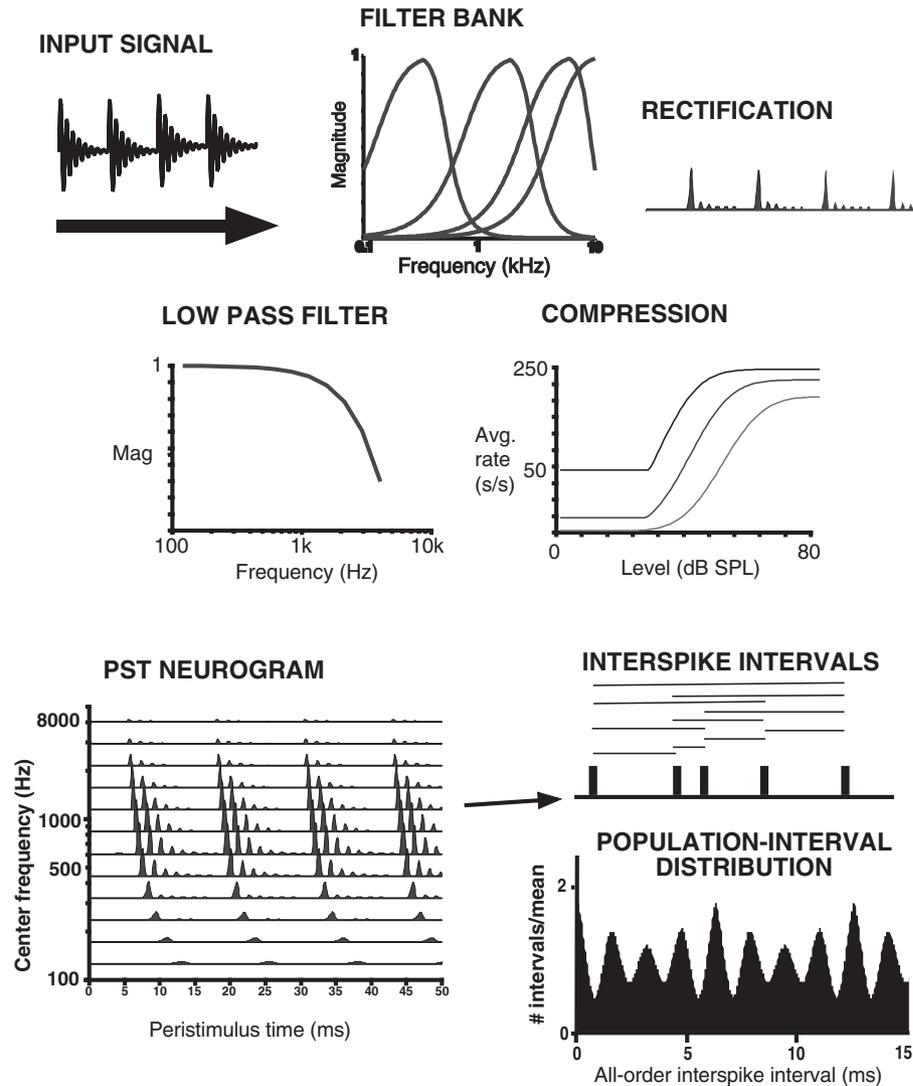


Fig. 2. Auditory nerve array simulation for the estimation of population-interval distributions. An input signal is passed through a bank of bandpass filters, half-wave rectified, low pass filtered, and compressed using three rate-level functions to produce post-stimulus time (PST) histograms for each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

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

Pitch frequency can be explicitly estimated by finding prominent peaks in population-interval distributions or by examining the repetition pattern of the whole histogram. Earlier estimations involved locating the first major peak in the interval distribution (Cariani & Delgutte, 1996a; Cariani & Delgutte, 1996b; Meddis & Hewitt, 1991a). More recently,

we have devised a more satisfying method for estimating pitch that takes into account repeating structure in the whole interval pattern. In this method, all intervals that are part of an interval series are counted, and the pitch is estimated to correspond to the series with the most intervals (highest mean bincount). For example, the sieve corresponding to 200 Hz contains intervals near 5, 10, 15, 20, 25, and 30 ms. This method is more general than peak-picking and is relevant to estimating the relative strengths of multiple pitches that can be produced by multiple interval subpatterns. The relative strength of a given pitch is estimated to be the ratio of the mean bincounts for its sieve to the mean of the whole distribution. The interval sieve is used in this context as an analysis of the all-order interval representations rather than as a hypothetical neural operation. A time-domain theory of pitch multiplicity and of pitch fusion can be built up from such

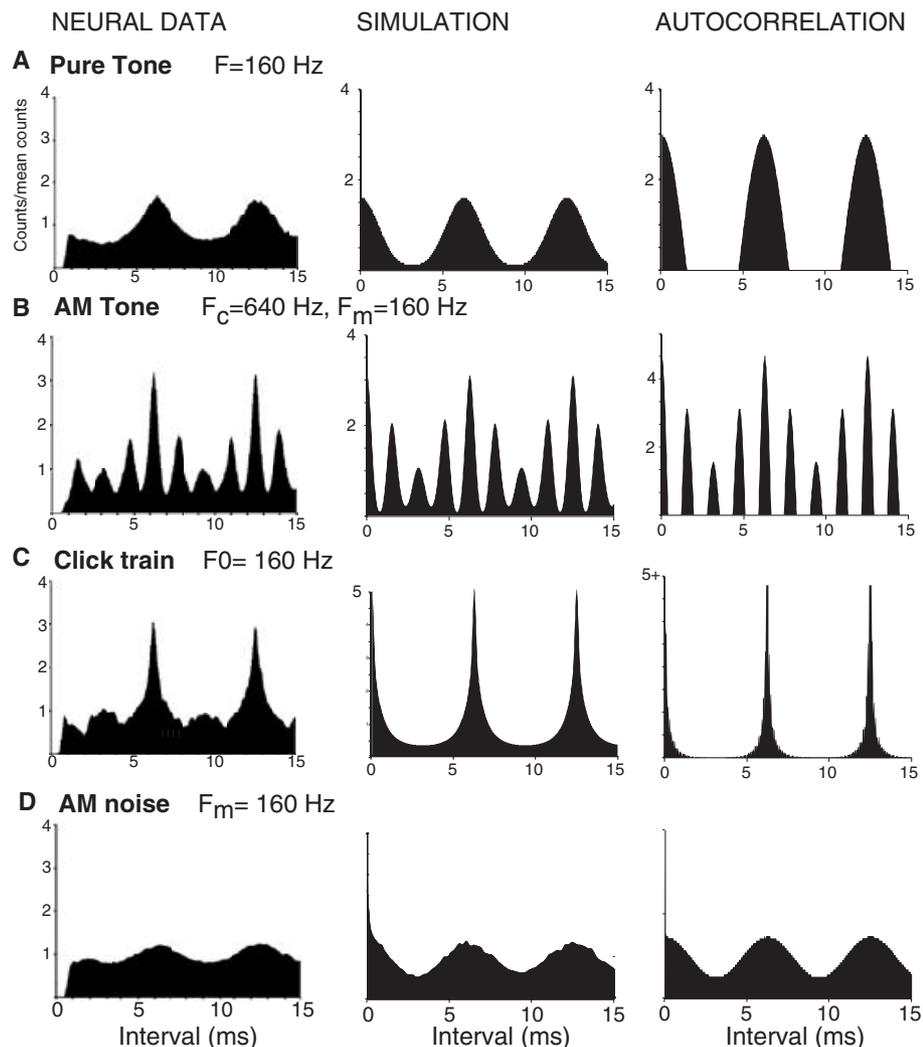


Fig. 3. Comparisons of population-interval distributions and autocorrelation function for six stimuli that produce a low pitch at 160 Hz. Left. Population interval distributions estimated from recorded responses of 50–100 auditory nerve fibers in Dial-anesthetized cats (Cariani & Delgutte, 1996). Middle.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

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

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

In the interval distributions and autocorrelations, those stimuli that produce strong pitches produce high peak-to-mean ratios in population-interval distributions ($p/m > 1.5$),

which means that a larger fraction of the intervals that they produce are pitch-related (e.g., at $1/F_0$ and its multiples). Those stimuli that produce weaker pitches produce lower peak-to-mean ratios ($1.3 < p/m < 1.5$), and those stimuli that fail to produce definite pitches produce ratios close to unity ($p/m < 1.3$).

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

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

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

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

these situations are far removed from those encountered in musical contexts.

Common temporal patterns and pitch similarity

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

Simulated population interval distributions for a series of fundamental frequencies related by different frequency ratios, including many found in a just-tempered scale are shown in Figure 4. These distributions have all been normalized to their means. Some simple relations are apparent. For both pure and complex tones, the distributions have common major peaks when ratios between fundamentals are near 2:1, 3:1, 3:2, and 4:3. These correspond to musical intervals of octaves, twelfths, fifths, and fourths. Distributions for $F_0 = 100$ (1:1), 200 (2:1), and 300 (3:1) share intervals at 10 and 20 ms. Distributions for $F_0 = 100$ and 150 Hz (3:2) share intervals at 20 ms, those for 133 and 200 Hz at 15 ms, those for 200 and 300 Hz at 10 and 20 ms. Distributions for $F_0 = 200$ and 167 Hz (4:3) share intervals at 20 ms. Fundamental ratios near these values, such as those produced by equal temperament tunings, also produce similar interval overlaps.

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

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

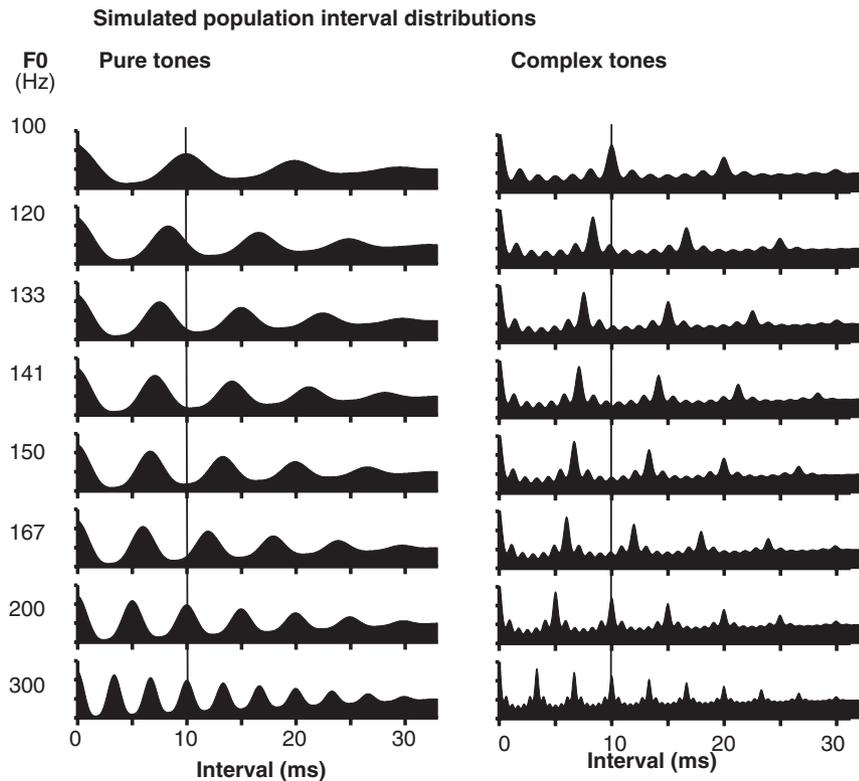


Fig. 4. Similarities between population-interval representations associated with different fundamental frequencies. Simulated population-interval distributions for pure tones (left) and complex tones (right) consisting of harmonics 1–6.

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

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

Population interval distributions for pure and complex tones are systematically compared in Figure 5. Pearson correlation coefficients (r) between all pairs of simulated population interval distributions associated frequencies from 30–440 Hz are plotted in the upper panel (A). For pure tones

(left correlation map) the highest correlations (darkest bands) follow unisons, octaves, and twelfths. For complex tones (right correlation map) there are also additional, fainter bands associated with fifths, fourths, and sixths. Cross sections of the two correlation maps are shown in the bottom panel, where the relative correlation strengths of all frequency ratios can be seen for a few selected notes.

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

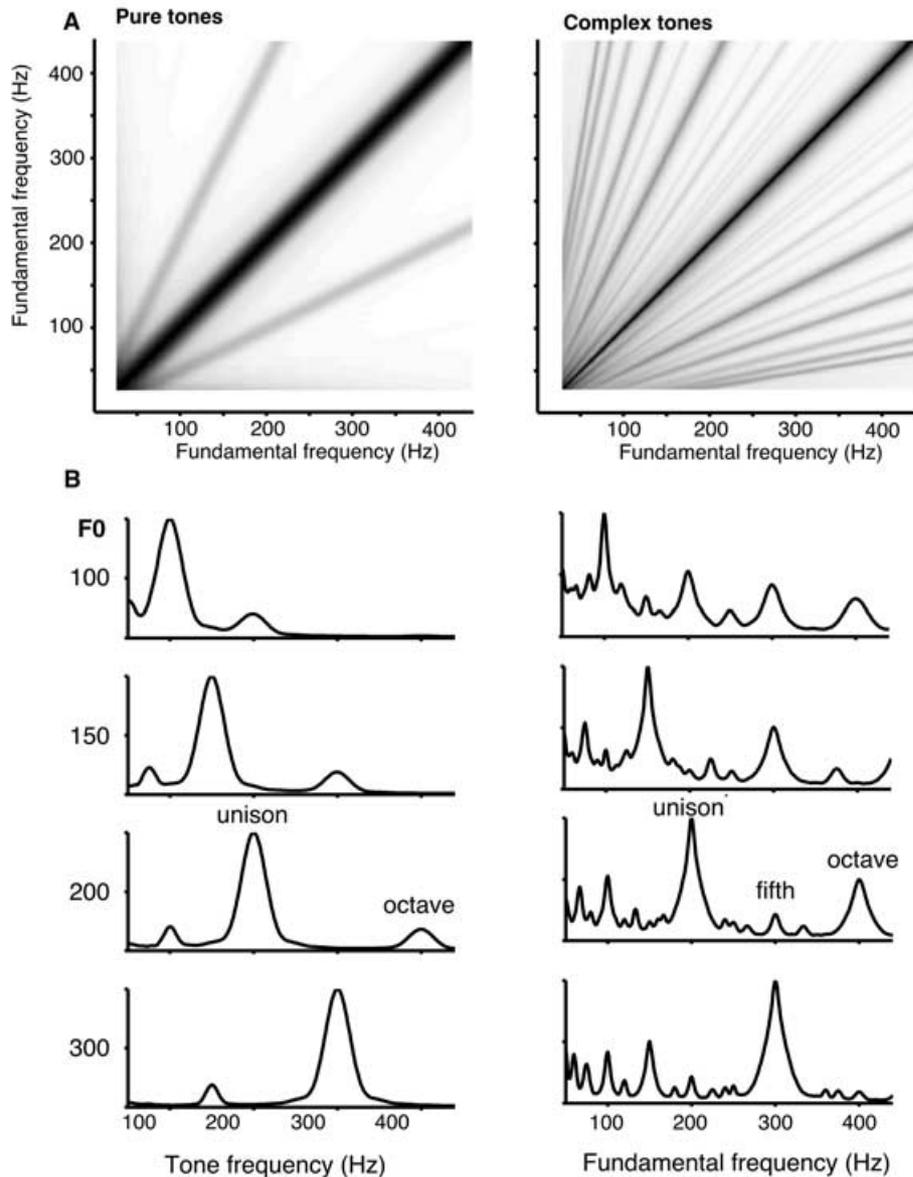


Fig. 5. Tonal structure and pattern similarities between population-interval distributions. Top. Map of correlation coefficients between all pairs of simulated population-interval distributions produced by pure and complex tones with fundamentals ranging from 1–440 Hz. B. Cross-sectional correlation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

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

in the microstructure of the neural codes and computations that subserve perception.

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

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

Temporal patterns and note-key relations

One can also assess similarities between interval patterns produced by individual notes and musical chords, and compare these to patterns of similarity judgments by human listeners (Handel, 1989; Krumhansl, 1990; Leman & Carreras, 1997). In a series of studies on tonal context, Krumhansl and colleagues developed a “probe tone” tech-

nique for investigating note-note and note-key relationships. In order to minimize possible effects of pitch height, they used notes and note triads made up of octave harmonics in the range from 80–2000 Hz. Notes were constructed in an equally-tempered chromatic scale. Key contexts were established by presenting scales followed by a major or minor triad followed by a probe tone. Experimenters then asked musically experienced listeners to judge how well a particular note “fit with” the previously presented chord. Their averaged, scaled “probe tone ratings” for C major and C minor key profiles are presented in the top left plots of Figure 6 (Krumhansl, 1991, p. 31). Similar judgements are also obtained using other stimuli and key-contexts, so these note-key relationships appear to be general in that they do not depend on particular, familiar key contexts.

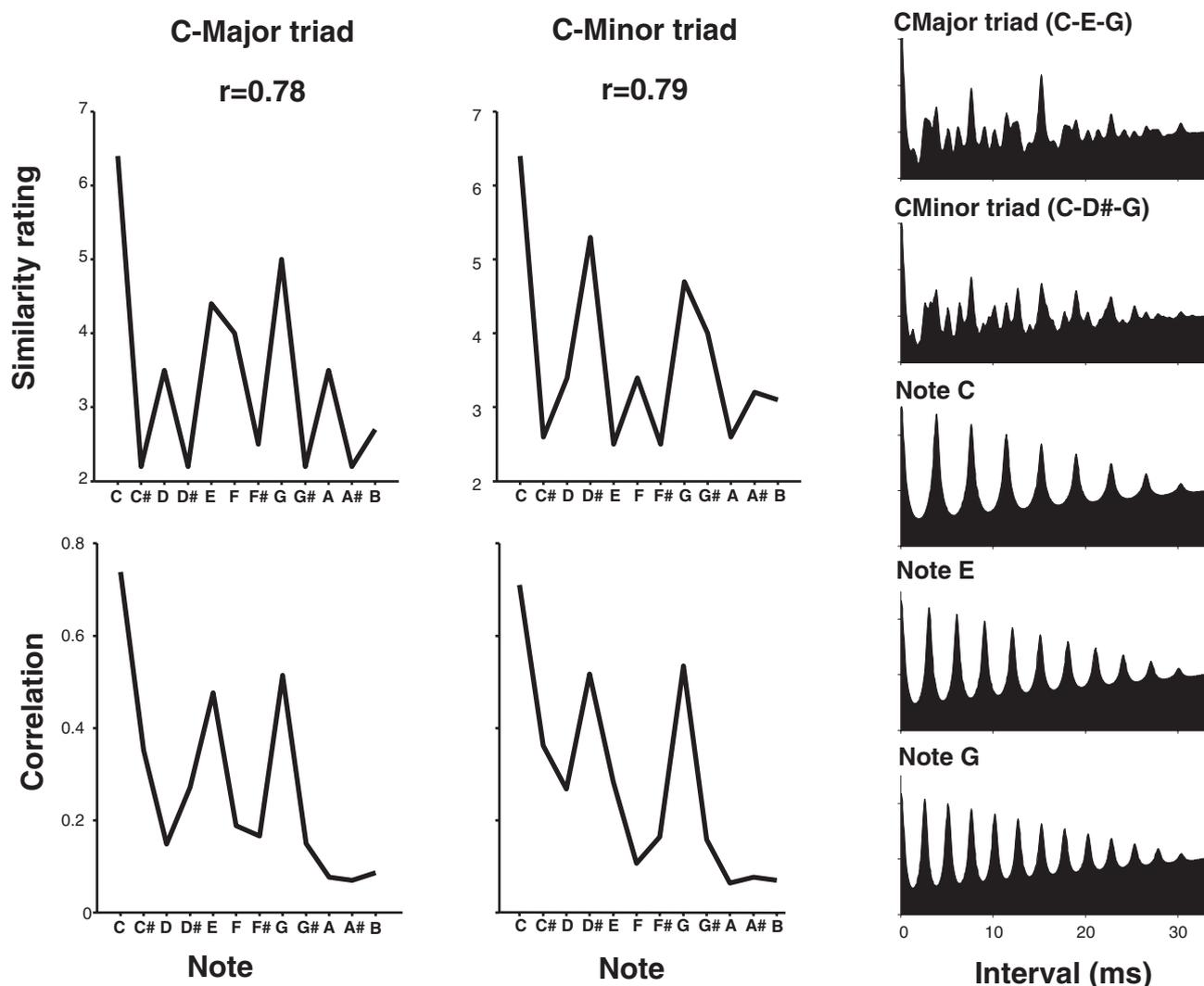


Fig. 6. Comparison of interval-based measures of note-chord similarity with human judgments. Left top. Results of probe tone experiments: human ratings of how well a note fits in with a preceding chord (Krumhansl, 1990). Chords were either C-major (CEG) or C-minor (CD#G) note triads. Notes consisted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

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

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

Overtones and undertones in autocorrelation-like representations

The weighted population-interval histograms of the two chords and two individual notes are shown in the panels on the right. The roots of chords, like the low pitches of harmonic complexes, produce patterns of major peaks in autocorrelation-like representations. These kinds of temporal representations seamlessly handle both harmonic and inharmonic patterns. Note that C-major and C-minor have major peaks at 7.6 ms, the period of their common root, C3 (132 Hz). Each chord pattern contains the interval patterns of its constituent notes. Each note pattern in turn is approximately the superposition of the intervals of each of its constituent harmonics. Because the autocorrelation of any periodic pattern contains intervals associated not only with the pattern's repetition, but also those associated with multiple repetition periods, the autocorrelation function contains subharmonics of each frequency component, and by super-

position, subharmonics of fundamental frequencies. The same arguments also apply to population-interval distributions (PID's). For example, a pure tone at 1000 Hz produces many all-order interspike intervals at 1 ms and its multiples, such that the interval peaks are located at 1, 2, 3, . . . ms lags. The note PID's in Figure 6 show peaks at fundamental periods and their subharmonics. In this sense autocorrelation and population-interval representations contain both overtones (harmonics) of musical sounds, because they are present in the acoustics, and their undertones (subharmonics), because they are periodic. A few explanations for the roots of chords based on undertone series have been raised in the past (Makeig, 1982), including Terhardt's algorithm for inferring virtual pitch from the subharmonics of frequency components (Terhardt, 1979). Although schemes based on subharmonics have been dismissed by music theorists (Hindemith, 1945) on grounds that they have no apparent representation in auditory frequency maps, clearly such a representation can be found in patterns of longer interspike intervals.

Implications for musical tonality

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

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

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

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

Neural timing nets

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

Rationale for development

By far the dominant assumption in both neuroscience and music theory is that the auditory system consists of an array of band-pass filters in the cochlea that produce spatial activation patterns in auditory frequency maps that are subsequently analyzed by connectionist networks (Bharucha, 1991; Bharucha, 1999; Cohen et al., 1994). While it is possible to arrange inter-element connectivities in a manner that permits the pitches of complex tones and their equivalence classes to be computed, in order for these networks to attain discriminative precisions on par with those of humans and animals, their inputs must be highly frequency selective and robust. With a few exceptions, the narrow tunings that are required are generally at odds with those that are seen in the auditory pathway, where neural response areas typically broaden greatly at moderate to high sound pressure levels.

Many modelers simply sidestep the issue by using very narrow frequency tunings that are derived from human psychophysical experiments, but this assumes away the mechanisms by which cochlear and neural responses produce fine discriminations in the first place. In the midst of frequency-domain operations on “central spectra” derived from psychophysically-derived auditory filters, it can easily be forgotten that the central spectra themselves may be based on interspike interval information rather than rate-place profiles (Goldstein & Sruлович, 1977; Moore, 1997a).

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

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

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

Table 1. General types of neural networks.

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

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

Types of neural networks

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

Traditionally, neural networks have been conceptualized in terms of spatialized activation patterns and scalar signals. Conventional connectionist nets generally assume synchro-

nous inputs whose time structure is generally irrelevant to the encoding of information. Whatever relevant temporal structure exists is converted to spatial activation patterns by means of temporal pattern detectors.

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

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

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

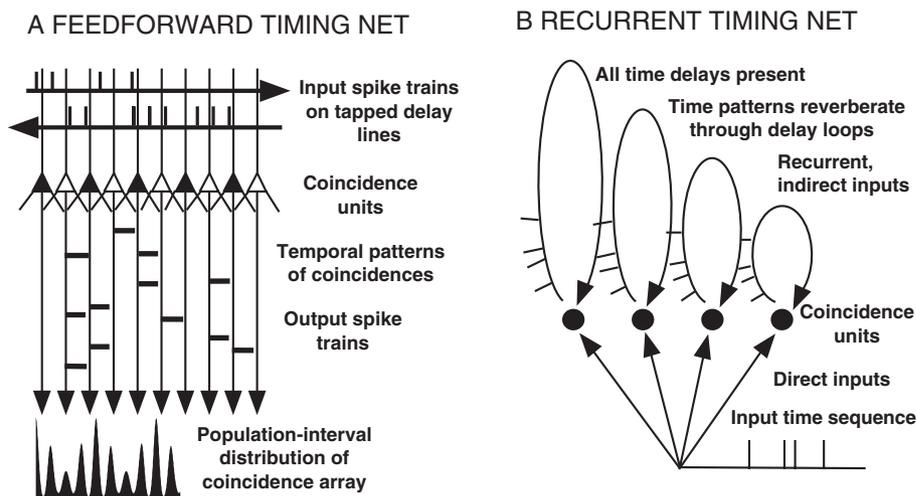


Fig. 7. Neural timing nets. (a) A simple feedforward timing net consisting of two tapped delay lines and a linear array of coincidence detectors. Outputs of coincidence detectors contain only temporal patterns that are common to the two inputs. The population interspike interval distribution of the outputs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

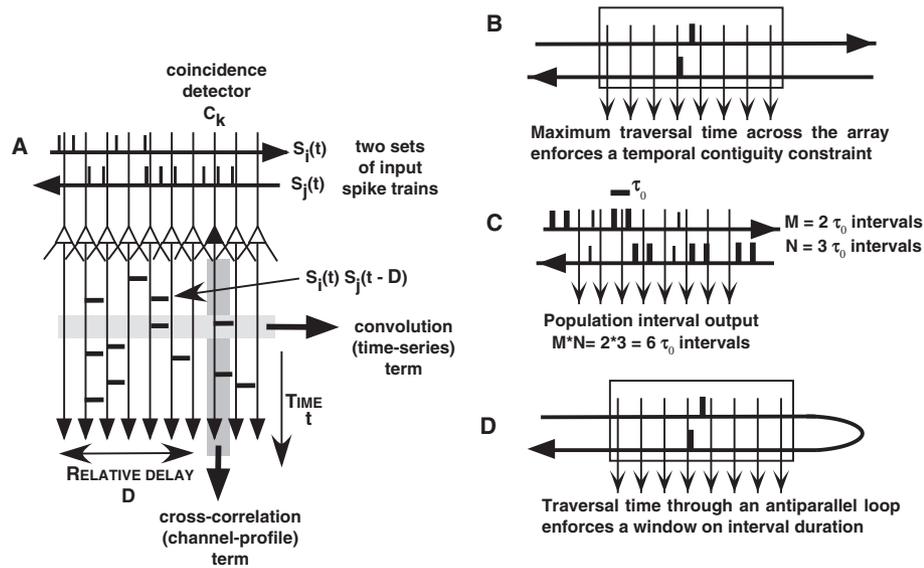


Fig. 8. A simple feedforward timing net. (a) General schematic of a coincidence array traversed by tapped delay lines. Summation over time in each output channel yields the cross-correlation function, while summation over output channels for each time yields the convolution of the two inputs. (b) The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

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

Timing networks were directly inspired by several temporal processing architectures. Feedforward timing networks are related to the Jeffress temporal cross-correlation architecture for binaural localization (Jeffress, 1948), Licklider's temporal auto-correlation architecture for pitch (Licklider, 1951; Licklider, 1959), the combination auto- and cross-correlation architectures of Licklider and Cherry (Cherry, 1961; Licklider, 1959), Braitenberg's cerebellar timing model (Braitenberg, 1961), and the temporal correlation memory strategies suggested by Longuet-Higgins (Longuet-Higgins, 1987; Longuet-Higgins, 1989). Although feedforward networks of all kinds have been studied in greater depth because of their formal tractability, in view of the ubiquity of reciprocal connectivities between neurons and neural populations, theoretical neuroscientists have always looked to recurrent networks as more realistic brain models. Thus, adaptive resonance circuits, re-entrant loops, and thalamocortical resonances are prominent in the current thinking about large scale neural integration. While conceptions incorporate notions of reverberating circuits (Hebb, 1949), these schemes use spatial patterns (Grossberg, 1988) and sequences of neural activations (e.g., McCulloch, 1969), rather than temporal codes. Nonetheless there have been a few proposals for temporal processing using

neural delay loops (Thatcher & John, 1977). In the auditory system, Patterson's strobed temporal integration model (Patterson, Allerhand, & Giguere, 1995) functions in a manner similar to a recurrent timing network in that it retains previous temporal patterns that are then cross-correlated with incoming ones to build up stable auditory images. The timing networks that we describe here are not adaptive networks that adjust inter-element connectivities (synaptic weights) or delays (conduction times), but such adaptive mechanisms have been proposed in the past (MacKay, 1962), and are important to any general hypothesis concerned with neurocomputational substrates of learning and memory.

Feed-forward timing networks

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

Basic computational properties

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

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

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

The traversal time across the array determines which parts of the signals interact with each other (Fig. 8b). All intervals from each set of inputs that arrive within the temporal con-

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

Extraction of common pitch and timbre

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

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

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

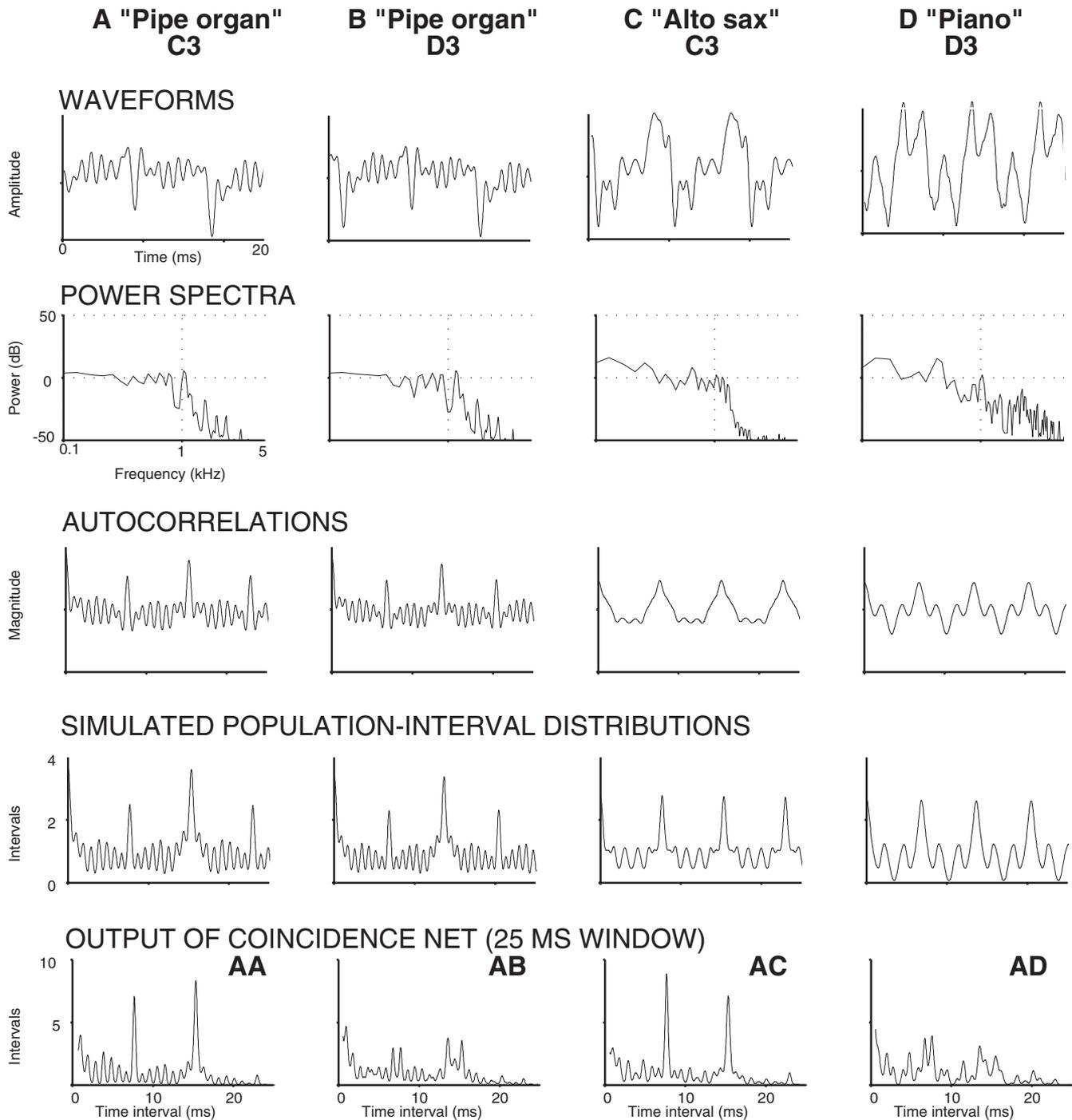


Fig. 9. Pitch matching irrespective of timbral difference. Waveforms, power spectra, autocorrelations, and simulated population-interval distributions are shown for four synthesized waveforms (Yamaha PSR-76). Complex tones A and B have similar spectral envelopes and produce the same timbre. Tones A and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

mental at 131Hz or 7.6ms, while D3 has a fundamental at 147Hz or 6.8ms. The correlation coefficients between the four population interval distributions are given in Table 2. Correlation comparisons between population interval distributions heavily weight commonalities of fundamental frequency (pitch) over those of spectral shape

(timbre). Correlations between the patterns in the different interval ranges associated with timbre (0–2 ms) and pitch (2–10 ms) of these stimuli are presented in Table 3. The “pipe organ” and “alto sax” voices have very similar short-interval patterns, but these differ substantially from that of the “piano” voice.

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

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

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

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

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

The feedforward timing network thus produces an interval distribution in its output that corresponds to the common pitch of the two inputs, and it does this without ever having to explicitly compute the fundamental frequency. This kind of comparison operation produces relative pitch judgements rather than the absolute ones that would be generated from an explicit pitch estimate. The relativistic nature of this periodicity-matching process parallels the relativity of pitch perception. In order to match pitches using this scheme, one adjusts the fundamental frequencies (as in tuning an instrument) so as to maximize the relative numbers of intervals produced by the coincidence array. Here the relative numbers of intervals produced by the stimulus combinations were AA (20), AB (16), AC (18), and AD (15), producing a similarity ordering of 1) common pitch and timbre, 2) common pitch,

slightly different timbre, 3) different pitch, slightly different timbre, and 4) different pitch and timbre. This simple strategy of pitch matching by maximizing the output of the whole array works as long as sound pressure levels (and consequently input spike rates) are kept constant, but would be expected to fail if matching were carried out with roving levels. Maximizing the peak to background ratio of intervals in the output that are associated with the pitch periodicity to be matched achieves a normalization operation that makes the computation more like a correlation comparison (as in Figs 5 and 6). Further development of the computation of similarity in timing nets should include incorporation of inhibitory elements that impose penalties for anticoincidences and perform cancellation-like operations (de Cheveigné, 1998; Seneff, 1985; Seneff, 1988).

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

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

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

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

Recurrent timing nets

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

Basic properties of recurrent timing nets

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

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

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

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

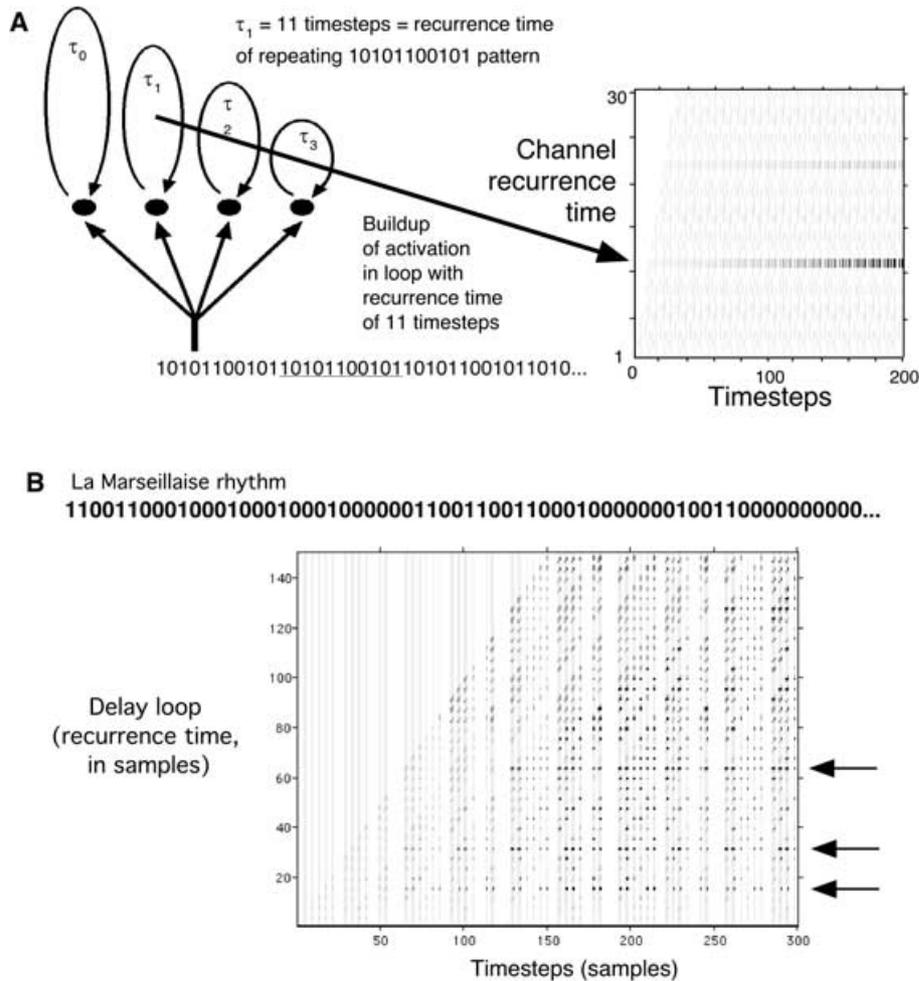


Fig. 10. Behavior of recurrent timing nets. (a) Behavior of a simple recurrent timing net for periodic pulse train patterns. The delay loop whose recurrence time equals the period of the pattern builds up that pattern. (b) Response of a recurrent timing net to the beat pattern of *La Marseillaise*. Arrows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

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

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

Many random, periodically repeating binary sequences were presented to the network. The network behavior for an input pulse train that repeats the same 11-step sequence (... 10101100101...) is shown in Figure 10a. Here coincidence windows are one timestep (sampling period). The plot on the right shows the value of the circulating pattern for

each delay loop as a function of time. The evolution of the circulating pattern can be interpreted as the build up of a perceptual image. The network builds up the 11-step sequence first in the loop whose recurrence time is 11 steps, and later in the loop whose recurrence time is 22 time steps. This behavior is entirely understandable, since it is in the delay loop whose recurrence time equals the period of the repeating pattern that the previous pattern maximally coincides with its repetition. If pulses are randomly added to the pattern (noise), then their pulse patterns do not reliably recur at periodic intervals, and consequently do not build up in any one delay loop. The network detects periodic patterns and enhances them relative to aperiodic ones.

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

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

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

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

Recurrent timing nets for computing rhythmic expectation

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

One of the shortcomings of the simple 5% buildup rule discussed above is that given a periodic signal, the response pattern builds up geometrically, and this dramatically favors shorter periodicities over longer ones. In order to rectify this imbalance, the buildup factor B that regulates the rate of increase of the loop signal was adjusted in proportion to the loop delay LD_k , i.e., $B_k = LD_k/100$, such that longer delay loops have proportionately higher facilitative factors. This equalizes in a crude way shorter and longer loops, which have differences in the number of times the signals are subjected to coincidence and facilitation per unit time. Subsequent applications of these networks to the problem of

separating concurrent vowels have used buildup rules that saturate more gracefully, where the output of a given coincidence unit is the minimum of direct and circulating inputs plus some fraction of their difference. The rule that describes the coincidence operation was $S_k(t) = \min(S_{\text{direct}}(t), B * S_{\text{direct}}(t) * S_{\text{loop}}(t))$, where $S_k(t)$ is the output of coincidence element k associated with delay loop of recurrence time LD_k , $S_{\text{direct}}(t)$ is the incoming direct input signal, and $S_{\text{loop}}(t) = S_k(t - LD_k)$.

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

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

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

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

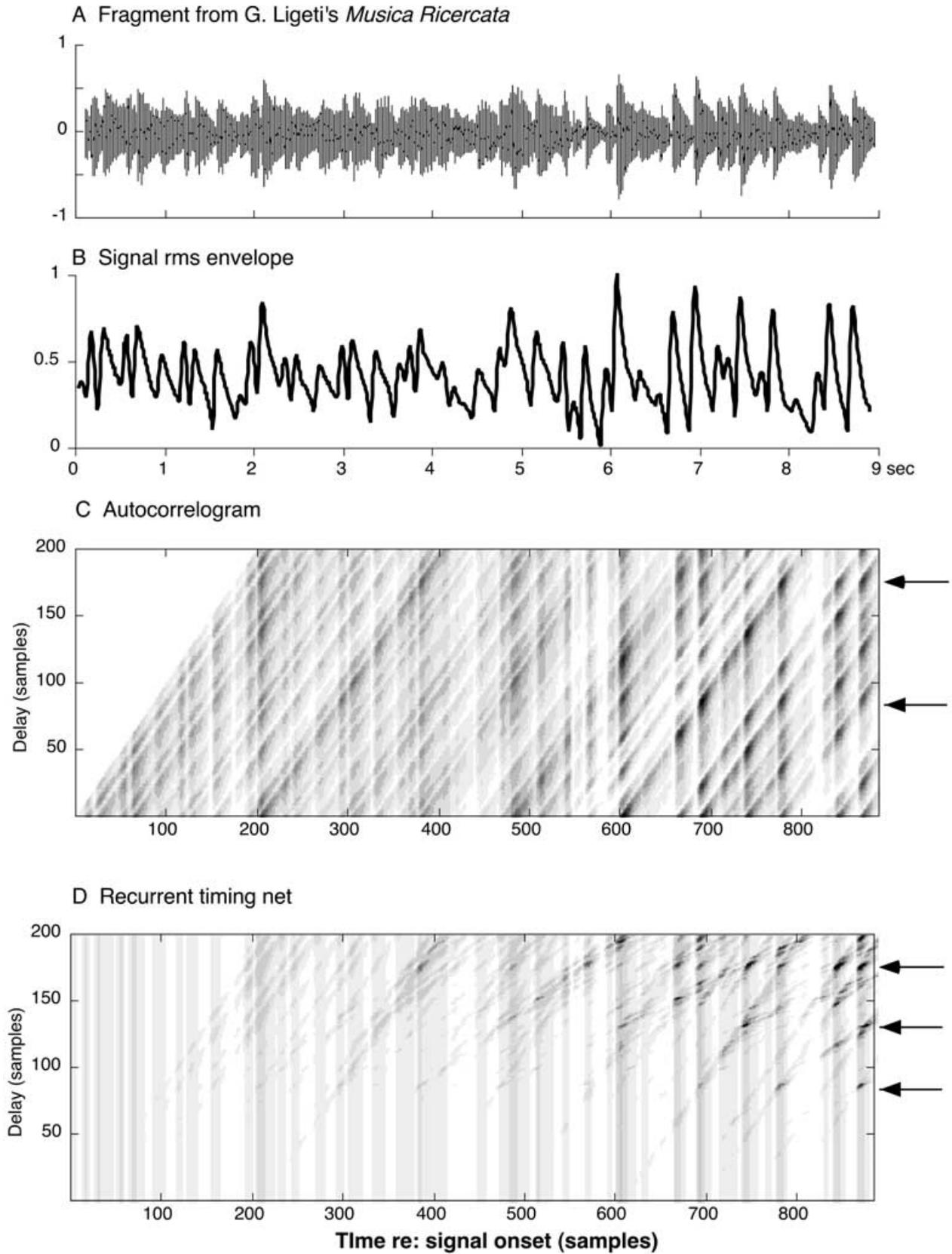


Fig. 11. Analysis of complex rhythmic pattern in music. (a) Waveform fragment from Ligeti, *Musica Ricercata*. (b) Rms envelope of the waveform. (c) Autocorrelogram (running autocorrelation) of the envelope. (d) Response of the recurrent timing net. Arrows indicate delay channels that built up prominent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

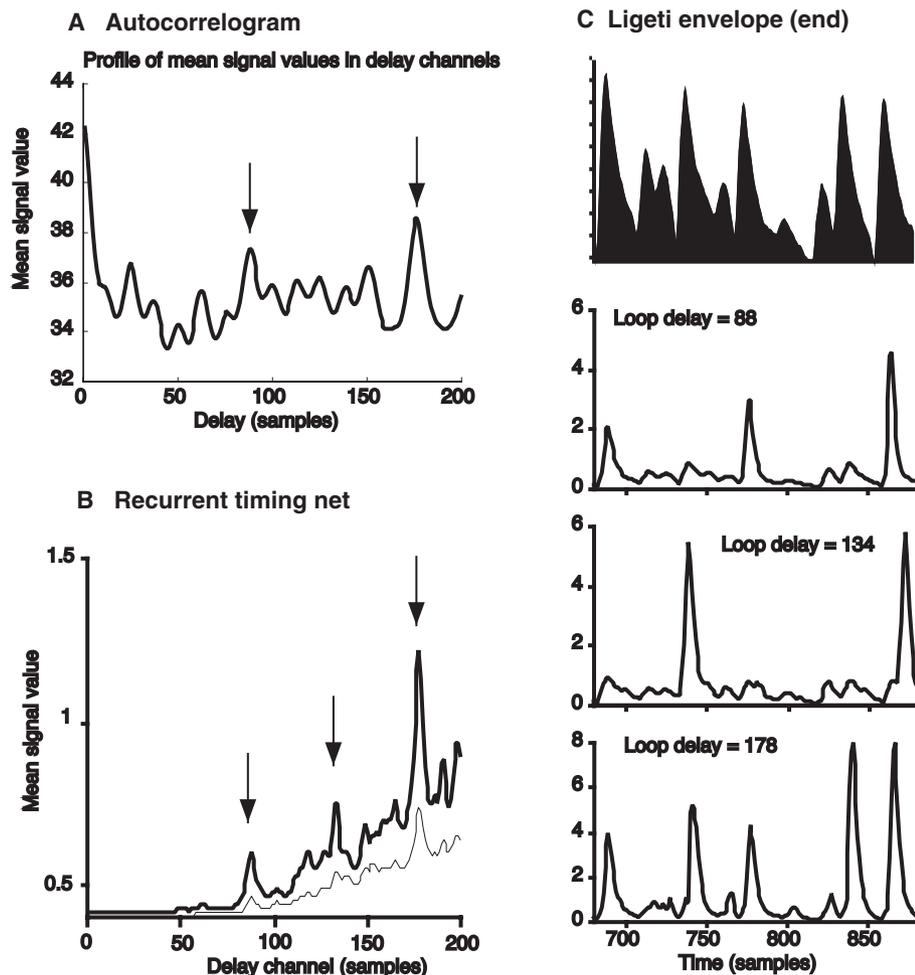


Fig. 12. Analysis of Ligeti fragment by autocorelograms and timing net. (a) Average autocorrelation value as a function of delay (mean value of autocorrelogram). (b) Average amplitude of signals in timing net loops as a function of loop delay. Thicker line shows average signal over the last 200 samples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

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

The response of the recurrent net to the Ligeti fragment is similarly shown in Figure 11d and in Figures 12b and c. The recurrent net builds up the same periodicities that are seen in the running autocorrelation: at 88 and 178 samples duration, plus an additional one at 134 samples (1.34 s).

These can also be seen in the peaks in the mean signal values in each delay loop that are plotted in Figure 12b. The additional peak is due to the difference between the non-iterated autocorrelation of the autocorrelogram and the exponential nature of the iterated pattern-buildup process of the recurrent net. The upward trend in the mean signal strength profile is due to the loop-dependent scaling of the buildup factor that was discussed above.

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

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

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

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

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

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

terns using feedforward networks that are largely indifferent to phase.

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

Conclusions

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

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

References

- Bharucha, J.J. (1991). Pitch, harmony and neural nets: a psychological perspective. In: P. Todd & G. Loy (Eds.), *Connectionism and Music* (pp. 84–99). Cambridge: MIT Press.
- Bharucha, J.J. (1999). Neural nets, temporal composites, and tonality. In: D. Deutsch (Ed.) (pp. 413–440). San Diego: Academic Press.
- Boomsalter, P. & Creel, W. (1962). The long pattern hypothesis in harmony and hearing. *Journal of Music Theory*, 5, 2–31.
- Boring, E.G. (1942). *Sensation and Perception in the History of Experimental Psychology*. New York: Appleton-Century-Crofts.
- Braitenberg, V. (1961). Functional interpretation of cerebellar histology. *Nature*, 190, 539–540.
- Brugge, J.F., Anderson, D.J., Hind, J.E., & Rose, J.E. (1969). Time structure of discharges in single auditory nerve fibers of the squirrel monkey in response to complex periodic sounds. *Journal of Neurophysiology*, 32, 386–401.
- Cariani, P. (1995). As if time really mattered: temporal strategies for neural coding of sensory information. *Communication and Cognition – Artificial Intelligence (CC-AI)*, 12(1–2), 161–229. Reprinted in: K. Pribram (Ed.), *Origins: Brain and Self-Organization*, Hillsdale, NJ: Lawrence Erlbaum, 1994; 1208–1252.
- Cariani, P. (1999). Temporal coding of periodicity pitch in the auditory system: an overview. *Neural Plasticity*, 6, 147–172.
- Cariani, P. (2001a). Neural timing nets for auditory computation. In: S. Greenberg & M. Slaney (Eds.), *Computational Models of Auditory Function* (pp. 235–249). Amsterdam: IOS Press.
- Cariani, P. (2001b). Symbols and dynamics in the brain. *Biosystems*, 60, 59–83.
- Cariani, P. (2001c). Temporal coding of sensory information in the brain. *Acoust. Sci. & Tech.*, 22, 77–84.
- Cariani, P. (2001d). Neural timing nets. *Neural Networks*, 14, 737–753.
- Cariani, P., Delgutte, B., & Tramo, M. (1997). Neural representation of pitch through autocorrelation. *Proceedings, Audio Engineering Society Meeting (AES), New York, September 1997, Preprint #4583 (L-3)*.
- Cariani, P.A. & Delgutte, B. (1996a). Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *Journal of Neurophysiology*, 76, 1698–1716.
- Cariani, P.A. & Delgutte, B. (1996b). Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase-invariance, pitch circularity, and the dominance region for pitch. *Journal of Neurophysiology*, 76, 1717–1734.
- Cherry, C. (1961). Two ears – but one world. In: W.A. Rosenblith (Ed.), *Sensory Communication* (pp. 99–117). New York: MIT Press/John Wiley.
- Chistovich, L.A. (1985). Central auditory processing of peripheral vowel spectra. *Journal of the Acoustical Society of America*, 77, 789–805.
- Chistovich, L.A. & Malinnikova, T.G. (1984). Processing and accumulation of spectrum shape information over the vowel duration. *Speech Communication*, 3, 361–370.
- Clarke, E.F. (1999). Rhythm and timing in music. In: D. Deutsch (Ed.) (pp. 473–500). San Diego: Academic Press.
- Clynes, M. & Walker, J. (1982). Neurobiologic functions, rhythm, time, and pulse in music. In: M. Clynes (Ed.), *Music, Mind, and Brain: the Neuropsychology of Music* (pp. 171–216). New York: Plenum.
- Cohen, M.A., Grossberg, S., & Wyse, L.L. (1994). A spectral network model of pitch perception. *J. Acoust. Soc. Am.*, 98, 862–879.
- de Boer, E. (1956). *On the “residue” in hearing*. Unpublished Ph.D., University of Amsterdam.
- de Boer, E. (1976). On the “residue” and auditory pitch perception. In: W.D. Keidel & W.D. Neff (Eds.), *Handbook of Sensory Physiology* (Vol. 3, pp. 479–583). Berlin: Springer Verlag.
- de Cheveigné, A. (1998). Cancellation model of pitch perception. *Journal of the Acoustical Society of America*, 103, 1261–1271.
- Delgutte, B. (1996). Physiological models for basic auditory percepts. In: H. Hawkins, T. McMullin, A.N. Popper, & R.R. Fay (Eds.), *Auditory Computation* (pp. 157–220). New York: Springer Verlag.
- Demany, L. & Semal, C. (1988). Dichotic fusion of two tones one octave apart: evidence for internal octave templates. *Journal of the Acoustical Society of America*, 83, 687–695.
- Demany, L. & Semal, C. (1990). Harmonic and melodic octave templates. *Journal of the Acoustical Society of America*, 88, 2126–2135.
- DeWitt, L.A. & Crowder, R.G. (1987). Tonal fusion of consonant musical intervals: The oomph in Stumpf. *Perception and Psychophysics*, 41, 73–84.
- Epstein, D. (1995). *Shaping Time: Music, Brain, and Performance*. New York: Simon & Schuster Macmillan.
- Essens, P.J. & Povel, D.-J. (1985). Metrical and nonmetrical representations of temporal patterns. *Perception and Psychophysics*, 37, 1–7.
- Evans, E.F. (1978). Place and time coding of frequency in the peripheral auditory system: some physiological pros and cons. *Audiology*, 17, 369–420.
- Fay, R.R. (1988). *Hearing in vertebrates: a psychophysics data-book*. Winnetka, Ill.: Hill-Fay Associates.

- Fraisse, P. (1978). Time and rhythm perception. In: E.C. Carterette & M.P. Friedman (Eds.), *Handbook of Perception. Volume VIII. Perceptual Coding* (pp. 203–254). New York: Academic Press.
- Goldstein, J.L. (1973). An optimum processor theory for the central formation of the pitch of complex tones. *Journal of the Acoustical Society of America*, 54, 1496–1516.
- Goldstein, J.L. & Sruлович, P. (1977). Auditory-nerve spike intervals as an adequate basis for aural frequency measurement. In: E.F. Evans & J.P. Wilson (Eds.), *Psychophysics and Physiology of Hearing* (pp. 337–346). London: Academic Press.
- Gray, P.M., Krause, B., Atema, J., Payne, R., Krumhansl, C., & Baptista, L. (2001). The music of nature and the nature of music. *Science*, 291, 52–54.
- Greenberg, S. (1980). *Neural Temporal Coding of Pitch and Vowel Quality: Human Frequency-Following Response Studies of Complex Signals*. Los Angeles: UCLA Working Papers in Phonetics #52.
- Grossberg, S. (1988). *The Adaptive Brain, Vols I. and II*. New York: Elsevier.
- Hall III, J.W. & Peters, R.W. (1981). Pitch for nonsimultaneous successive harmonics in quiet and noise. *Journal of the Acoustical Society of America*, 69, 509–513.
- Handel, S. (1989). *Listening*. Cambridge: MIT Press.
- Hebb, D.O. (1949). *The Organization of Behavior*. New York: Simon & Schuster.
- Hindemith, P. (1945). *The Craft of Musical Composition. I. Theoretical Part* (Arthur Mendel, Trans.). New York: Associated Music Publishers.
- Hunt, F.V. (1978). *Origins in Acoustics*. Woodbury, NY: Acoustical Society of America.
- Jeffress, L.A. (1948). A place theory of sound localization. *J. Comp. Physiol. Psychol.*, 41, 35–39.
- John, E.R. (1967). Electrophysiological studies of conditioning. In: G.C. Quarton, T. Melnechuk, & F.O. Schmitt (Eds.), *The Neurosciences: A Study Program* (pp. 690–704). New York: Rockefeller University Press.
- John, E.R. (1972). Switchboard vs. statistical theories of learning and memory. *Science*, 177, 850–864.
- Jones, M.R. (1976). Time, our lost dimension: toward a new theory of perception, attention, and memory. *Psychological Review*, 83, 323–255.
- Jones, M.R. (1978). Auditory patterns: studies in the perception of structure. In: E.C. Carterette & M.P. Friedman (Eds.), *Handbook of Perception. Volume VIII. Perceptual Coding* (pp. 255–288). New York: Academic Press.
- Jones, M.R. (1987). Perspectives on musical time. In: A. Gabriellson (Ed.), *Action and Perception in Rhythm and Music* (pp. 153–175). Stockholm: Royal Swedish Academy of Music.
- Jones, M.R. & Hahn, J. (1986). Invariants in sound. In: V. McCabe & G.J. Balzano (Eds.), *Event Cognition: An Ecological Perspective* (pp. 197–215). Hillsdale, New Jersey: Lawrence Erlbaum.
- Kaernbach, C. & Demany, L. (1998). Psychophysical evidence against the autocorrelation theory of auditory temporal processing. *Journal of the Acoustical Society of America*, 104, 2298–2306.
- Keidel, W. (1984). The sensory detection of vibrations. In: W.W. Dawson & J.M. Enoch (Eds.), *Foundations of Sensory Science* (pp. 465–512). Berlin: Springer-Verlag.
- Keidel, W.D. (1992). The phenomenon of hearing: an interdisciplinary discussion. II. *Naturwissenschaften*, 79, 347–357.
- Kiang, N.Y.S., Watanabe, T., Thomas, E.C., & Clark, L.F. (1965). *Discharge Patterns of Single Fibers in the Cat's Auditory Nerve*. Cambridge: MIT Press.
- Krumhansl, C.L. (1990). *Cognitive Foundations of Musical Pitch*. New York: Oxford University Press.
- Langner, G. (1983). Evidence for neuronal periodicity detection in the auditory system of the Guinea Fowl: implications for pitch analysis in the time domain. *Experimental Brain Research*, 52, 33–355.
- Langner, G. (1992). Periodicity coding in the auditory system. *Hearing Research*, 60, 115–142.
- Large, E.W. (1994). *Dynamic representation of musical structure*. Unpublished Ph.D., The Ohio State University.
- Leman, M. (1999). Time domain filter model of tonal induction. *Tonality Induction, Proceedings of the Ninth FWO Research Society on Foundations of Music Research, University of Ghent, April 6–9, 1999*, 53–87.
- Leman, M. (2000). An auditory model of the role of short-term memory in probe-tone ratings. *Music Perception*, 17, 481–510.
- Leman, M. & Carreras, F. (1997). Schema and Gestalt: Testing the hypothesis of psychoneural isomorphism by computer simulation. In: M. Leman (Ed.), *Music, Gestalt, and Computing* (pp. 144–165). Berlin: Springer.
- Leman, M. & Schneider, A. (1997). Origin and nature of cognitive and systematic musicology: an introduction. In: M. Leman (Ed.), *Music, Gestalt, and Computing* (pp. 11–29). Berlin: Springer.
- Leman, M. & Verbeke, B. (2000). The concept of minimal ‘energy’ change (MEC) in relation to Fourier transform, auto-correlation, wavelets, AMDF, and brain-like timing networks – application to the recognition of repetitive rhythmical patterns in acoustic musical signals. In: K. Jokien, D. Helylen, & A. Nijholt (Eds.), *Proceedings, Workshop on Internalizing Knowledge (Nov. 22–24, 2000, Ieper, Belgium)* (pp. 191–200). Ieper, Belgium: Cele-Twente.
- Licklider, J.C.R. (1951). A duplex theory of pitch perception. *Experientia*, VII(4), 128–134.
- Licklider, J.C.R. (1954). “Periodicity” pitch and “place” pitch. *Journal of the Acoustical Society of America*, 26, 945.
- Licklider, J.C.R. (1956). Auditory frequency analysis. In: C. Cherry (Ed.), *Information Theory* (pp. 253–268). London: Butterworth.
- Licklider, J.C.R. (1959). Three auditory theories. In: S. Koch (Ed.), *Psychology: A Study of a Science. Study I. Conceptual and Systematic* (Vol. Volume I. Sensory, Perceptual, and Physiological Formulations, pp. 41–144). New York: McGraw-Hill.

- Longuet-Higgins, H.C. (1987). *Mental processes: Studies in cognitive science*. Cambridge, Mass.: The MIT Press.
- Longuet-Higgins, H.C. (1989). A mechanism for the storage of temporal correlations. In: R. Durbin, C. Miall, & G. Mitchison (Eds.), *The computing neuron* (pp. 99–104). Wokingham, England: Addison-Wesley.
- Lyon, R. & Shamma, S. (1996). Auditory representations of timbre and pitch. In: H. Hawkins, T. McMullin, A.N. Popper, & R.R. Fay (Eds.), *Auditory computation* (pp. 221–270). New York: Springer Verlag.
- Mach, E. (1898). *Popular scientific lectures, Third Edition*. La Salle, Illinois: Open Court.
- MacKay, D.M. (1962). Self-organization in the time domain. In: M.C. Yovitts, G.T. Jacobi, & G.D. Goldstein (Eds.), *Self-Organizing Systems 1962* (pp. 37–48). Washington, D.C.: Spartan Books.
- Makeig, S. (1982). Affective versus analytic perception of musical intervals. In: M. Clynes (Ed.), *Music, mind, and brain: The neuropsychology of music* (pp. 227–250). New York: Plenum.
- McCulloch, W.S. (1947). Modes of functional organization of the cerebral cortex. *Federation Proceedings*, 6, 448–452.
- McCulloch, W.S. (1969). Of digital oscillators. In: K.N. Leibovic (Ed.), *Information processing in the nervous system* (pp. 293–296). New York: Springer Verlag.
- McKinney, M. (1999). A possible neurophysiological basis of the octave enlargement effect. *Journal of the Acoustical Society of America*, 106, 2679–2692.
- Meddis, R. & Hewitt, M.J. (1991a). Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I. Pitch identification. *Journal of the Acoustical Society of America*, 89, 2866–2882.
- Meddis, R. & Hewitt, M.J. (1991b). Virtual pitch and phase sensitivity of a computer model of the auditory periphery. II. Phase sensitivity. *Journal of the Acoustical Society of America*, 89, 2883–2894.
- Meddis, R. & O'Mard, L. (1997). A unitary model of pitch perception. *Journal of the Acoustical Society of America*, 102, 1811–1820.
- Meyer, L.B. (1956). *Emotion and meaning in music*. Chicago: University of Chicago.
- Moelants, D. (1997). A framework for the subsymbolic description of meter. In: M. Leman (Ed.), *Music, Gestalt, and Computing* (pp. 263–276). Berlin: Springer.
- Moore, B.C.J. (1997a). *Introduction to the psychology of hearing* (Fourth ed.). London: Academic Press.
- Moore, B.C.J. (1997b). *Introduction to the psychology of hearing* (Fourth ed.). London: Academic Press.
- Morrell, F. (1967). Electrical signs of sensory coding. In: G.C. Quarton, T. Melnechuck, & F.O. Schmitt (Eds.), *The neurosciences: A study program* (pp. 452–469). New York: Rockefeller University Press.
- Ohgushi, K. (1983). The origin of tonality and a possible explanation for the octave enlargement phenomenon. *Journal of the Acoustical Society of America*, 73, 1694–1700.
- Orbach, J. (1998). *The neuropsychological theories of Lashley and Hebb*. Lanham: University Press of America.
- Palmer, A.R. (1992). Segregation of the responses to paired vowels in the auditory nerve of the guinea pig using autocorrelation. In: M.E.H. Schouten (Ed.), *The auditory processing of speech* (pp. 115–124). Berlin: Mouton de Gruyter.
- Parncutt, R. (1989). *Harmony: A psychoacoustical approach*. Berlin: Springer-Verlag.
- Patterson, R.D. (1986). Spiral detection of periodicity and the spiral form of musical scales. *Psychology of Music*, 14, 44–61.
- Patterson, R.D. (1994). The sound of a sinusoid: time-interval models. *Journal of the Acoustical Society of America*, 96, 1560–1586.
- Patterson, R.D., Allerhand, M.H., & Giguere, C. (1995). Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform. *Journal of the Acoustical Society of America*, 98, 1890–1894.
- Perkell, D.H. & Bullock, T.H. (1968). Neural coding. *Neurosciences Research Program Bulletin*, 6, 221–348.
- Pressnitzer, D., Patterson, R.D., & Krumboltz, K. (2001). The lower limit of melodic pitch. *Journal of the Acoustical Society of America*, 109, 2074–2084.
- Rieke, F., Warland, D., de Ruyter van Steveninck, R., & Bialek, W. (1997). *Spikes: Exploring the neural code*. Cambridge: MIT Press.
- Rose, J.E. (1980). Neural correlates of some psychoacoustical experiences. In: D. McFadden (Ed.), *Neural mechanisms of behavior* (pp. 1–33). New York: Springer Verlag.
- Scheirer, E.D. (1998). Tempo and beat analysis of acoustic musical signals. *Journal of the Acoustical Society of America*, 103, 588–601.
- Scheirer, E.D. (2000). *Music listening systems*. Cambridge, Mass.: Ph.D. Dissertation, M.I.T.
- Schneider, A. (1997). “Verschmelzung”, tonal fusion, and consonance: Carl Stumpf revisited. In: M. Leman (Ed.), *Music, Gestalt, and Computing*. Berlin: Springer.
- Schneider, A. (2001, in press). Inharmonic sounds: implications as to “pitch”, “timbre” and “consonance”. *Journal of New Music Research*.
- Schouten, J.F. (1940). The residue, a new concept in subjective sound. *Proceedings Koninklijke Nederlandse van Wetenschappen*, 43, 356–365.
- Schouten, J.F., Ritsma, R.J., & Cardozo, B.L. (1962). Pitch of the residue. *Journal of the Acoustical Society of America*, 34, 1418–1424.
- Schreiner, C.E. & Langner, G. (1988). Coding of temporal patterns in the central auditory system. In: G.M. Edelman, W.E. Gall, & W.M. Cowan (Eds.), *Auditory function: Neurobiological bases of hearing* (pp. 337–362). Wiley: New York.
- Schwarz, D.W.F. & Tomlinson, R.W.W. (1990). Spectral response patterns of auditory cortical neurons to harmonic complex tones in alert monkey (*Macaca mulatta*). *Journal of Neurophysiology*, 64, 282–298.
- Seneff, S. (1985). *Pitch and spectral analysis of speech based on an auditory synchrony model*. Unpublished Ph.D., M.I.T.

- Seneff, S. (1988). A joint synchrony/mean-rate model of auditory speech processing. *Journal of Phonetics*, 16, 55–76.
- Sethares, W.A. (1999). *Tuning, timbre, spectrum, scale*. London: Springer.
- Sethares, W.A. & Staley, T. (2001). Meter and periodicity in musical performance. *Journal of New Music Research*, 30, 149–158.
- Shamma, S. & Sutton, D. (2000). The case of the missing pitch templates: How harmonic templates emerge in the early auditory system. *Journal of the Acoustical Society of America*, 107, 2631–2644.
- Shepard, R.N. (1964). Circularity in judgments of relative pitch. *Journal of the Acoustical Society of America*, 36, 2346–2353.
- Siebert, W.M. (1968). Stimulus transformations in the peripheral auditory system. In: P.A. Kollers & M. Eden (Eds.), *Recognizing patterns* (pp. 104–133). Cambridge: MIT Press.
- Simmons, A.M. & Ferragamo, M. (1993). Periodicity extraction in the anuran auditory nerve. *J. Comp. Physiol. A*, 172, 57–69.
- Slaney, M. (1998). Connecting Correlograms to neurophysiology and psychoacoustics. In: A.R. Palmer, A. Rees, A.Q. Summerfield, & R. Meddis (Eds.), *Psychophysical and physiological advances in hearing* (pp. 563–569). London: Whurr.
- Slaney, M. & Lyon, R.F. (1993). On the importance of time – a temporal representation of sound. In: M. Cooke, S. Beet, & M. Crawford (Eds.), *Visual representations of speech signals* (pp. 95–118). New York: John Wiley.
- Terhardt, E. (1973). Pitch, consonance, and harmony. *Journal of the Acoustical Society of America*, 55, 1061–1069.
- Terhardt, E. (1979). Calculating virtual pitch. *Hearing Research*, 1, 155–182.
- Thatcher, R.W. & John, E.R. (1977). *Functional neuroscience, Vol. I. Foundations of cognitive processes*. Hillsdale, NJ: Lawrence Erlbaum.
- Tramo, M.J. (2001). Music of the hemispheres. *Science*, 291, 54–56.
- Tramo, M.J., Cariani, P.A. Delgutte, B., & Braida, L.D. (2001). Neurobiological foundations for the theory of harmony in Western tonal music. *Annals of the New York Academy of Sciences*, 930, 92–116.
- Troland, L.T. (1929a). *The principles of psychophysiology, Vols I–III*. New York: D. Van Nostrand.
- Troland, L.T. (1929b). The psychophysiology of auditory qualities and attributes. *Journal of Gen. Psychol.*, 2, 28–58.
- Turgeon, M. (1999). *Cross-spectral auditory grouping using the paradigm of rhythmic masking release (Ph.D. Thesis)*. Montreal: McGill University.
- Turgeon, M., Bregman, A.S., & Ahad, P.A. (in press). Rhythmic masking release: Contribution of cues for perceptual organization to the cross-spectral fusion of concurrent narrow-band noise. *Journal of the Acoustical Society of America*.
- van Noorden, L. (1982). Two channel pitch perception. In: M. Clynes (Ed.), *Music, mind and brain* (pp. 251–269). New York: Plenum.
- Weinberg, N.M. (1999). Music and the auditory system. In: D. Deutsch (Ed.) (pp. 47–112). San Diego: Academic Press.
- Wever, E.G. (1949). *Theory of hearing*. New York: Wiley.
- White, L.J. & Plack, C.J. (1998). Temporal processing of the pitch of complex tones. *Journal of the Acoustical Society of America*, 103(4), 2051–2059.
- Wiegrebe, L. (2001). Searching for the time constant of neural pitch extraction. *Journal of the Acoustical Society of America*, 109(3), 1082–1091.

Temporal Codes and Computations for Sensory Representation and Scene Analysis

Peter A. Cariani

Abstract—This paper considers a space of possible temporal codes, surveys neurophysiological and psychological evidence for their use in nervous systems, and presents examples of neural timing networks that operate in the time-domain. Sensory qualities can be encoded temporally by means of two broad strategies: stimulus-driven temporal correlations (phase-locking) and stimulus-triggering of endogenous temporal response patterns. Evidence for stimulus-related spike timing patterns exists in nearly every sensory modality, and such information can be potentially utilized for representation of stimulus qualities, localization of sources, and perceptual grouping. Multiple strategies for temporal (time, frequency, and code-division) multiplexing of information for transmission and grouping are outlined. Using delays and multiplications (coincidences), neural timing networks perform time-domain signal processing operations to compare, extract and separate temporal patterns. Separation of synthetic double vowels by a recurrent neural timing network is used to illustrate how coherences in temporal fine structure can be exploited to build up and separate periodic signals with different fundamentals. Timing nets constitute a time-domain scene analysis strategy based on temporal pattern invariance rather than feature-based labeling, segregation and binding of channels. Further potential implications of temporal codes and computations for new kinds of neural networks are explored.

Index Terms—Autocorrelation, multiplexing, neural networks, neural timing nets, temporal coding.

I. INTRODUCTION

THE project of understanding how the brain works as an information processing device is a exercise in reverse-engineering: we seek to understand the functional principles by which an highly effective system of unknown design achieves its goals. In order to proceed, we need some grasp of its functional capabilities, its internal structural organization, the input–output behavior of its processing elements, the means by which internal signals represent information, and the nature of information-processing operations that are carried out on those signals [1]. Once functional principles are understood, artificial devices can be designed that rival and exceed their biological counterparts.

The neural coding problem involves identification of dimensions of neuronal activity that play particular functional roles: which aspects subserve which informational functions. We will consider basic types of temporal pulse codes and how they convey information. We define a neural code by its functional role in the system: which changes in neural activity constitute differences in information content that are used by the rest of

the system to guide behavior [2], [3]. Our aim is to consider representations and computations that different temporal codes afford, rather than respective information transmission capacities [4] or underlying biophysical mechanisms [5]. Temporal codes can be used for representation of stimulus properties, localization of sources, and perceptual grouping.

In Sections II–V, various types of temporal codes are outlined, followed by examples from sensory neurophysiology. Section VI discusses temporal codes for signal multiplexing and perceptual grouping. Latter sections present examples of neural timing nets for temporal processing, and contemplate future possibilities.

II. TYPES OF TEMPORAL CODES

Pulse codes afford many different means of encoding information. Many catalogues and taxonomies of possible neural pulse codes have been proposed [2], [3], [6]–[13]. Although many are extensive, none covers all possible codes. Some examples are depicted in Fig. 1(A)–(K). The most basic distinction to be made is between channel-based codes and temporal codes.

Channel-based codes convey sensory distinctions through patterns of neural channel activations. The neural “labeled line” channel determines the nature of the information conveyed. In rate-based codes [Fig. 1(A)], stimulus properties are encoded via differences in neural discharge rates. “Rate-place” schemes encode patterns and property-combinations through across-neuron profiles of average discharge rates (which neurons fire how frequently). Rate-place representations constitute the neural coding assumption that has historically predominated in the neurosciences and in neural networks.

In contrast to channel-codes, temporal codes [Fig. 1(B)–(G), (I)] convey distinctions through relative timings of spikes such that alterations of spike timings change functional meanings of messages conveyed. Temporal codes range from coarse temporal modulations of discharge rate [14]–[16] to temporal correlations between individual spikes. Time codes can serve as vehicles for representation of stimulus attributes [2], [17], [18] or for their grouping into perceptual objects [19]–[24].

Temporal codes can be divided by whether they rely on recurring temporal patterns of spikes in the same channel(s) or on relative timings spikes arriving through different channels. The simplest temporal pattern code is an interspike interval code [Fig. 1(C)] in which time durations between spikes, produced by the same neuron(s), convey information about stimulus properties [17]. More complex temporal patterns can represent multiple properties via multiple interspersed intervals [Fig. 1(D)], [25], [26] and interval sequences [Fig. 1(E)], [18], [27], [28].

Manuscript received June 9, 2003; revised December 30, 2003.

The author is with the Eaton Peabody Laboratory, Massachusetts Eye and Ear Infirmary, Boston, MA 02114 USA (e-mail: cariani@epl.meei.harvard.edu).

Digital Object Identifier 10.1109/TNN.2004.833305

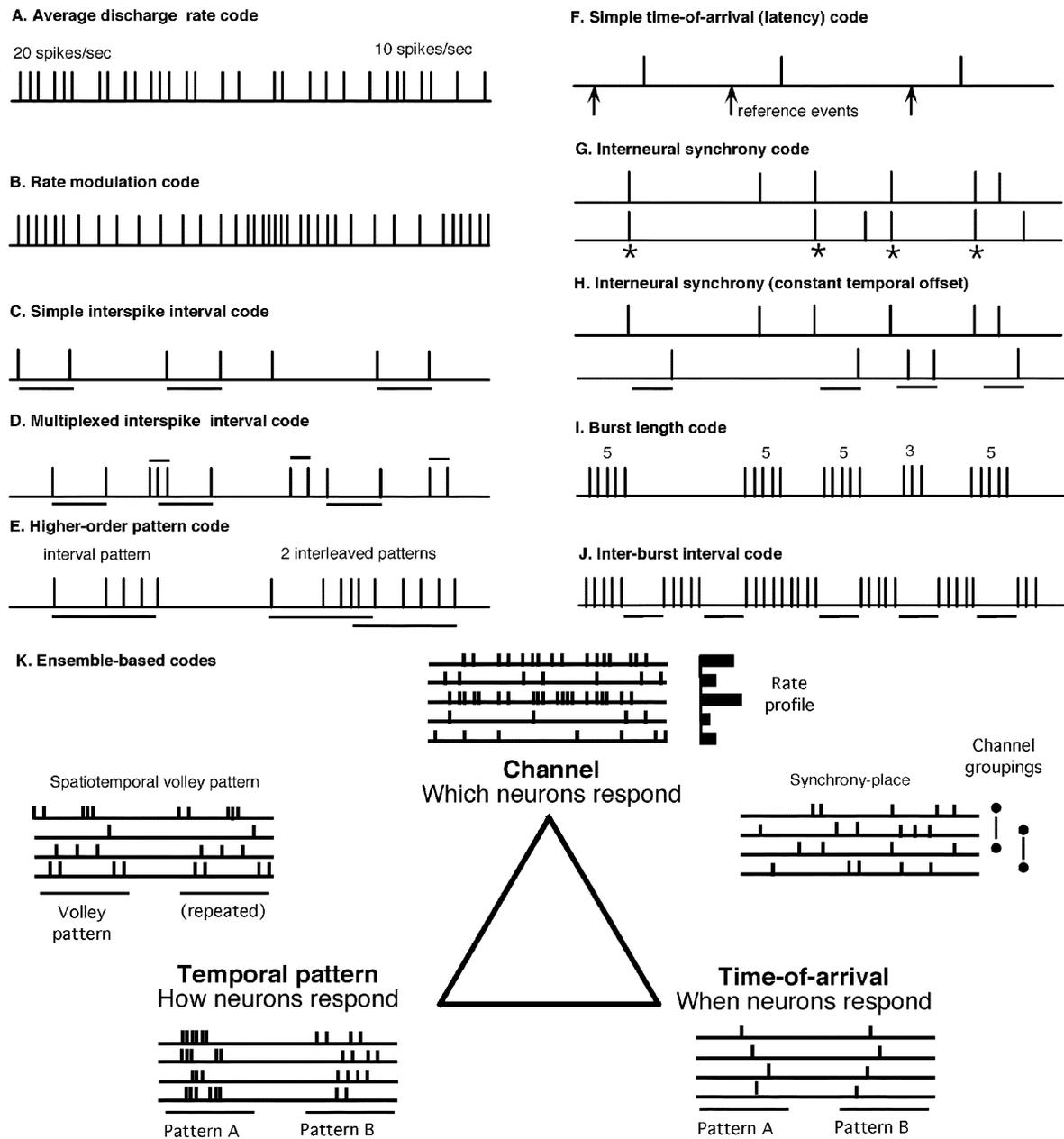


Fig. 1. Examples of neural pulse code schemes. (A) Rate-based channel code (# spikes in a given duration). (B) Rate modulation code (coarse temporal modulations of firing rate). (C) Simple interspike interval (ISI) code. (D) Multiplexing of two independent sets of intervals. (E) Complex temporal pattern code. (F) Simple time-of-arrival code (spike latency re: some reference event). (G) Interneuronal synchrony code (simultaneity). (H) Interneuronal synchrony (constant temporal offset). (I) Burst length code (nontemporal spike count code). (J) Interburst interval code (temporal pattern of burst events). (K) Combinations of neural ensemble-based channel, temporal pattern, and latency codes. Clockwise. Rate-channel scheme: rate profiles. Synchrony-place scheme: grouping of channels via common arrival time (spike synchrony). Time-of-arrival: two different latency-place patterns. Temporal pattern: two different ensemble-wide interval patterns. Spatiotemporal volley pattern (combination of all three code types): the same volley pattern shown twice.

Time-of-arrival or relative latency codes [Fig. 1(F)–(H)] use timing relations between spikes produced by different neurons. A well-known example [Fig. 1(H)] is found in binaural hearing, where time differences of sonic wavefronts reaching the two ears create different relative spike latencies in the two monaural auditory pathways. Here relative timing of spikes *across channels* rather than temporal structure *within channels* conveys information about source location. Since first spike latencies typically shorten with stimulus intensity [2], [29], [30], latency profiles (Fig. 1(K) lower right) can provide neuronal intensity cues. Time-of-arrival codes encompass interneuronal synchronies,

be they simultaneous [Fig. 1(G)] or delayed (Fig. 1(H), constant temporal offset).

While the taxonomy presented here is quite comprehensive in that it covers codes based on both spatial and temporal spike patterns, others stand outside its framework. For example, a burst length code [Fig. 1(I)] that counts spikes within a burst is neither a channel-code nor a temporal code. Similarly, an interburst interval code [Fig. 1(J)] relies on intervals between burst-events rather than between individual spikes. Still other schemes could rely on sequences of channel activations, spike times, and temporal patterns, i.e., ordinal rather than metrical

relations [31] Coarse temporal coding schemes intermediate between rate and spike correlation codes have also been proposed [10], [11]. Joint, statistical properties of neural ensembles, such as fractions of channels activated [32] or synchronized, variability of responses [2], [33], lie together with many other possible population-based codes [34], [35].

III. EXTRINSIC AND INTRINSIC TEMPORAL CODING MECHANISMS

Temporal codes are more common than is usually thought. Examination of the neurophysiology literature reveals stimulus-dependent temporal response patterns in early processing stages of virtually every sensory modality. Information-bearing temporal response patterns can be produced in two ways: extrinsically, through the temporal locking of neuronal discharges to stimulus waveforms (“phase-locking”), and intrinsically, through stimulus-specific triggering of endogenously generated time courses of response.

In many sensory systems, spike timings follow stimulus waveforms. Examples from auditory nerve fibers in the cat [36] and visual units in the monkey lateral geniculate nucleus [37] are shown in Fig. 2(A) and (B). Phase-locked responses also exist to mechanical vibrations of the skin [38], [39], to accelerations in the vestibular system, to inhalation cycles and odor pulses in olfaction [40], to self-produced electrical field oscillations and their external distortions in electroception [41], and to the movements of muscles via stretch receptors [42].

In the absence of stimulus locked responses, intrinsic temporal response patterns characteristic of particular stimuli, receptors, individual neurons, local neural circuits, and/or whole neural populations can bear information about the stimulus. Complex temporal response patterns are seen in the chemical senses. In early olfactory coding, many characteristic time courses of response depend upon both odorant and receptor [40], [43]–[45]. Time-delay neural nets for artificial noses [46] have been developed to effectively exploit relative latency patterns to discriminate odors. In later processing timings of discharge relative to oscillating field potentials are thought to play a role in organizing odor percepts [43]. In the rat gustatory system, different classes of tastants elicit characteristic temporal response patterns [Fig. 2(C)] that are associated with specific behavioral responses [47], [48]. Strikingly, playback of respective temporal patterns via electrical stimulation elicits appropriate behavioral responses in other rats, while scrambling of the patterns does not. In visual systems, intrinsic interspike interval patterns and coarse temporal modulations of firing rate result from differences in form, texture, and wavelength [14], [25], [49], [50]. The existence of subjective colors elicited by flicker-stimuli [51], [52] and temporally structured electrical stimulation [53] suggest the possibility of an intrinsic temporal code for color [12], [13].

IV. TIME-OF-ARRIVAL CODING OF STIMULUS LOCATION AND MOTION

Temporal patterns of activation across different sensory surfaces provide highly robust cues for stimulus direction. In audition, mechanoeception, and electroception, there may exist anal-

ogous mechanisms that make use of this cue to translate temporal differences into apparent location [2], [3], [41], [54], [55]. In all of these systems, receptors phase-lock to their respective adequate stimuli, such that the temporal structure of the stimulus is faithfully impressed on the timings of spikes produced by primary sensory neurons. By virtue of phase-locking, relative stimulus arrival times at different receptor sites are translated into relative spike latencies across sensory pathways. Relative spike timings can then in turn analyzed via neural delays and temporal coincidence detectors.

A well-known example is the localization of sounds in the azimuthal plane by means of interaural time differences (ITD) [56], [57]. Humans are able to distinguish differences of ITD as small as $\sim 20 \mu\text{s}$. Wavefronts from sound sources not directly in front of an observer arrive at the two ears at different times. These wavefronts produce phase-locked spikes in auditory nerve fibers whose relative timings reflect the interaural time differences. In the auditory brainstem, highly secure synapses, tapped delay lines, well-timed inhibitory inputs [5], and neural coincidence detectors in effect implement binaural cross-correlation operations that provide a readout of interaural time delays, and consequently, of azimuth estimates [41], [57], [58].

Analogous processing mechanisms may exist in mechanoeception, where relative delays of mechanical stimulation at different skin locations of a ms or more manifest themselves perceptually as changes in apparent location of stimulation [54]. As in audition, the apparent location moves toward the sensory surface that leads in time. Mechanoeceptive receptors phase-lock to skin deformations and hair displacements [38], [39], such that relative timings of neuronal discharges reflect relative timings of the stimulus arrival at different surfaces.

Active sensing involves observation of effects of actions on external environments. In bat and cetacean echolocation, acoustic signals are emitted and compared with echo patterns that contain time delays that provide information about distances and shapes of objects. In bat echolocation, time delays on the microsecond scale are registered by the relative timings between spikes produced by cries and their echoes [59], [60]. In some bat systems low-frequency modulations caused by interactions between the emitted signal and its Doppler shifted echo provide potential temporal cues for relative velocity (e.g., target flight speed and fluttering wingbeats) [59]. Another time-based active sensing strategy is found in electroception [41], [61]. Weakly electric fish produce sinusoidally-varying electrical fields around their bodies that are deformed by the presence of nearby external objects. These deformations alter the relative phases of the electric field at different body locations, which alter the relative latencies of spikes produced in afferent electroceptive pathways. As in the binaural example, these pathways have highly secure, low jitter connections, neural delays, and central coincidence detectors that permit use of sub-microsecond time-of-arrival differences.

Visual receptor arrays can be considered as collections of receptor surfaces. Phase-locking to temporal modulations of luminance produced by moving spatial patterns is ubiquitous in the visual systems of animals. As a consequence, temporal correlations between spikes produced in different visual channels

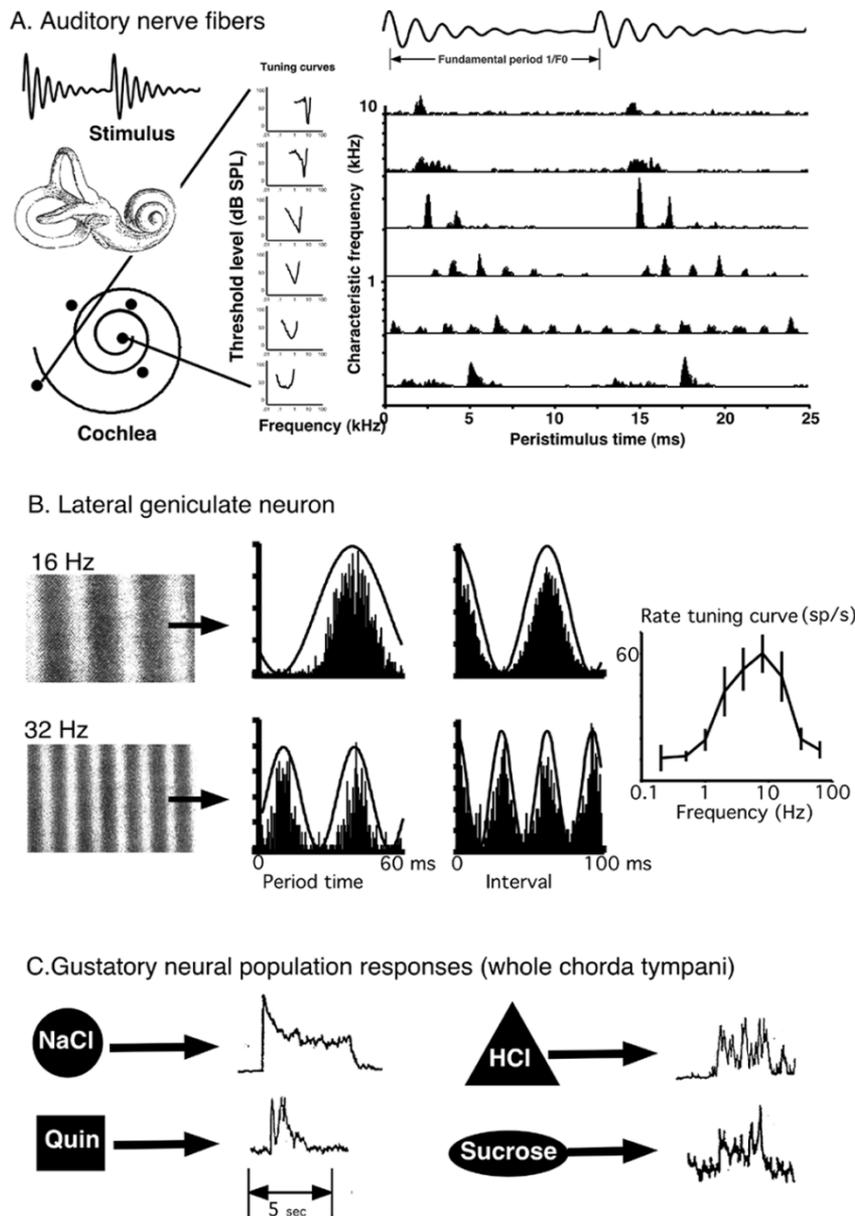


Fig. 2. Extrinsic and intrinsic temporal codes in early sensory processing. (A) Phase-locking of six cat auditory nerve fibers to a low-frequency harmonic complex tone. Rate-threshold tuning curves and poststimulus time histograms reflect effects of both cochlear filtering and the fine time structure of the stimulus. (B) Phase-locking of a visual neuron in the lateral geniculate nucleus of an anesthetized monkey in response to drifting gratings corresponding to sinusoidal temporal luminance modulations of 16 and 32 Hz. Period histograms and all-order interspike interval distributions are shown (16 Hz: 1233 spikes, 5897 intervals; 32 Hz: 495 spikes, 1102 intervals). Rate-based spatial tuning curve for the unit ($\text{mean} \pm \text{sd}$). Data courtesy of Przybylski & Pollen. (C) Intrinsic time courses of response in the gustatory system to four tastants of different classes: 0.1M NaCl (salty), 0.1M quinine (bitter), 0.1M HCl (sour), 0.5M sucrose (sweet). Waveforms are typical whole-nerve responses recorded from the chorda tympani of decerebrate rats. From Covey (1980).

potentially provide a general neurocomputational basis for the representation of visual motion. In the fly visual system, different spike timings in neighboring ommatidia are used for detection of motion [62]–[64]. Temporal crosscorrelation of small numbers of spikes permits rapid and precise motion estimations that inform flight course corrections in as little as 30 ms [63], [64].

Visual forms might also be encoded through temporal correlations between spikes produced across different retinal locations. Most vertebrate eyes are in constant drift even during fixation, and, many central visual neurons are known to discharge with relatively precise latency when contrast gradients (edges) cross their receptive fields. Moving images create

temporal modulations of luminance to which visual neurons phase-lock [Fig. 2(B)]. As in the auditory nerve, application of a temporally modulated stimulus converts a spike arrivals from a Poisson-like process to one dominated by the stimulus time structure. When sinusoidally-varying spatial gratings are drifted across receptive fields at constant velocities, at each retinal location spatial frequencies are converted to temporal frequencies of luminance modulation. Spike timings, as analyzed through peri-stimulus time (PST), period, and all-order interval histograms [Fig. 2(B)], faithfully replicate the temporal structure of the drifting image [65]. Temporal frequency can, thus, be accurately estimated from interspike interval statistics. For the lateral geniculate unit shown, temporal modulation

frequencies of 16 and 32 Hz were estimated to within 0.5% and 2%, respectively, of their true values on the basis of 5897 and 1102 all-order intervals. Given the coarseness of rate-based modulation tuning (rightmost plot), it is obvious that the spike timing information provides much finer precisions.

Similar stimulus-locking is seen for other visual stimuli. Many neurons in early stages of visual processing phase-lock to visual flickers of 50–100 Hz, but the relationship between perceptual flicker fusion limits and phase-locking is not simple [66]. Highest observed precisions of stimulus-driven spike timings in visual systems range from hundreds of microseconds to a few ms [63], [64], [67]–[69]. Stimulus-dependent synchronizations of spikes produced by different spatially remote neurons on the order of a few ms are also seen [70].

If the visual system used stimulus-driven spatial patterns of temporal correlations to register edges, then the disappearance of visual forms when an image is stabilized on the retina [71] could be readily explained. When images are stabilized, luminances are temporally unmodulated such that spike generation at each retinal position reverts to a Poisson-like process [72]. Temporal correlations of spikes across retinal positions disappear, as would spatial form information. The hypothesis is also consistent with recent psychophysical demonstrations that visual forms can be created through short-term spatiotemporal correlations that have no long-term spatial structure [73]. Such a hypothesis also potentially explains the hyperacuity problem in vision [64]—why visual acuity far exceeds the coarse spatial resolutions of individual receptive fields. It has been estimated from vernier acuity experiments with moving bars [74] that stimulus-driven spike timing with 1 ms jitter would be sufficient to account for observed performances, which lies within the realm of observed spike timings of visual units [75]. It also raises the possibility of spike-based spatiotemporal representations that couple to theories of visual form based on spatial auto and crosscorrelations [32], [76]–[78].

V. PHASE-LOCKED CODING OF STIMULUS WAVEFORMS

To the extent that sensory receptors follow the time structure of their adequate stimuli, that time structure is impressed on the timings of discharges of sensory neurons. As we have seen above, comparisons of the relative timings across receptors yields information about stimulus direction, movement, and perhaps, spatial form. For many kinds of stimuli, sensory qualities are related to the internal time structure of the stimulus waveform. Stimulus-locked temporal patterns of spikes in sensory neurons reflect this internal structure and, thus, provide information about the temporal form of the stimulus.

In the auditory system, phase-locked neural timing information is thought to be critical for most frequency discrimination [79]–[81]. Phase-locking of auditory nerve fibers to acoustic stimuli creates time intervals between spikes (interspike intervals) that are directly related to stimulus periodicities [Figs. 2(A) and 3(B)]. Distributions of interspike intervals from auditory neurons consequently contain information about the stimulus waveform and power spectrum for frequencies up to the limits of phase-locking. Phase-locking is maintained at virtually all sound pressure levels above response thresholds

such that interspike interval representations easily account for high precisions of auditory frequency discriminations over wide ranges of frequencies and sound pressure levels. As in binaural hearing, the time differences corresponding to the finest frequency discriminations (0.2% for pure tones near 1 kHz) are in the vicinity of 20 μ s. The progressively poorer ability of human listeners to distinguish higher frequencies from 2–10 kHz covaries with the progressive decline in phase-locking (and the quality of interval-based information) at those frequencies [79], [81]–[83]. Likewise, the frequency region for which there is good phase-locking, roughly up to 4 kHz, is coextensive with the existence region of musical tonality (octave matching, musical interval and melodic recognition).

Pitches of complex tones are explicable in similar terms. Harmonic complex tones produce pitches at their fundamentals (F0s), even in the absence of any spectral energy at that frequency (“pitch of the missing fundamental”). These low pitches of complex tones predominate in speech and music (they are the voice pitches we hear over telephones and cheap radios). In 1951, Licklider proposed a duplex time-delay neural network that operated on phase-locked auditory nerve responses to compute both rate-place and temporal autocorrelation representations of stimulus power spectra [81], [84], [85]. Decades later, the global interspike interval distribution of the auditory nerve was proposed as a candidate temporal representation for the pitch of complex tones. The general hypothesis has been confirmed in computer simulations [86] and in neurophysiological studies (Fig. 3) [36], [87]. With very few exceptions, the most common interval in the auditory nerve at any given time [Fig. 3(F)] corresponds to the pitch that is heard. On the basis of several thousand intervals, the fundamental pitch can be reliably estimated (within 1%). The fraction of pitch-related intervals *qualitatively* corresponds to the strength of the pitch that is heard. Such purely temporal, population-interval representations also account for many complex and subtle pitch phenomena: pitches of “missing fundamentals,” pitch equivalence classes, relative invariance of pitch and pitch salience with sound pressure level, pitches of stimuli having psychophysically-unresolved harmonics, the “nonspectral” pitches of amplitude modulated noise, as well as complex patterns of pitch shift that are heard for inharmonic stimuli. In effect, population-wide distributions of all-order interspike intervals form autocorrelation-like representations of the stimulus that in many (though not all) respects mirror stimulus autocorrelation functions. [87]; see [81] for recent developments and controversies. The close correspondences between the behavior of population-interval representations and psychophysically-observed patterns of pitch judgments provide very strong evidence that the interval information is utilized by the auditory system for pitch perception.

Purely temporal representations of vowel quality are also possible [80], [87]. Auditory nerve fibers phase-lock to frequency components in formant regions, consequently producing related intervals whose respective numbers depend on component intensities. Population-wide interval distributions consequently represent spectral energy distributions (formant patterns) through patterns of short intervals (0–4 ms) [85], [87]–[89]. These purely temporal representations of dominant

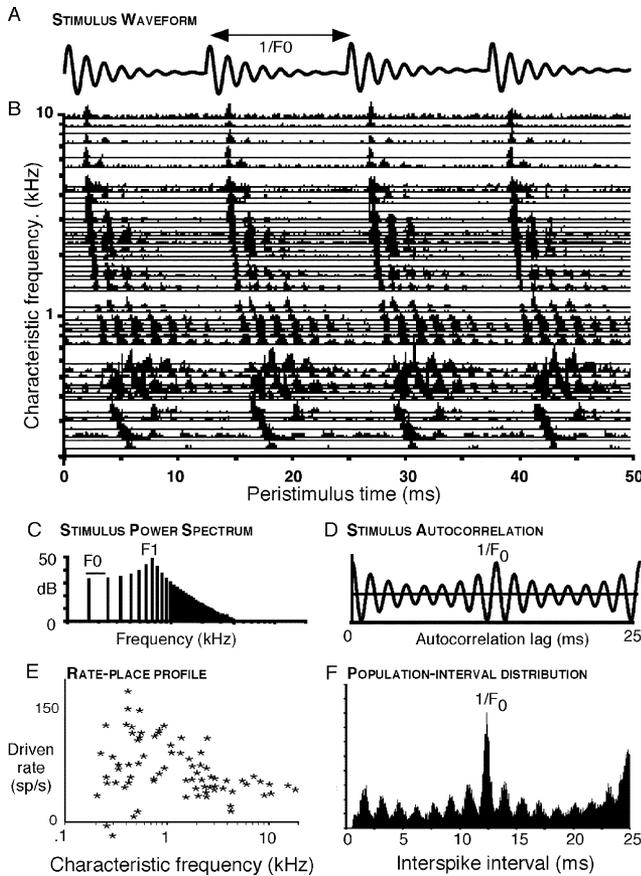


Fig. 3. Temporal coding of pitch in the auditory nerve. (A) Single-formant vowel waveform that produces a low pitch at its fundamental frequency ($F_0 = 80$ Hz). (B) Poststimulus time histograms of 52 cat auditory nerve fibers in response to 100 stimulus presentations at 60 dB SPL, arranged by characteristic frequency. (C) Stimulus power spectrum (log-frequency). (D) Stimulus autocorrelation function. (E) Population driven-rate profile (discharge rate—spontaneous rate). (F) Population-interval distribution (# all-order interspike intervals versus interval duration).

periodicity (pitch) and spectral energy distribution (timbre) have the high precision and robustness over large dynamic ranges that are characteristic of most auditory percepts. Their resilience to background noise [90] and sound separation capabilities [80] make them useful front-ends for artificial speech recognition systems.

The somatosensory analogue of auditory pitch is the sense of flutter-vibration [3], [38], [54]. Discrimination of frequencies up to 1 kHz appears to be based on interspike interval patterns produced through phase-locking of cutaneous afferents [39]. Complex tactile patterns produce corresponding temporal spike patterns [91] making an interval-based representations of tactile texture possible that parallel those for auditory pitch and timbre.

VI. TEMPORAL CODES IN SIGNAL MULTIPLEXING AND SCENE ANALYSIS

In addition to encoding sensory qualities themselves, temporal codes may be involved in perceptual grouping mechanisms. If object-properties are encoded via patterns of channel-activations, then the timing of those activations can be used to

sort them out. If object-properties themselves are encoded temporally, then commonalities of temporal pattern can be used as a basis for grouping.

Multiplexing involves the use of the same neural elements for concurrent representation and transmission of multiple signals. Multiplexing of information permits a given neuronal element to contribute to the representation of multiple perceptual objects, provided that there is a means of associating signals with particular objects. Three temporal strategies for binding ensembles of channels are time-division, frequency-division, and code-division multiplexing [Fig. 4(A)–(C)].

In time division multiplexing [Fig. 4(A)], the same channels participate in the encoding of multiple objects albeit at different times. If objects are represented by patterns of channels (e.g., feature detectors) then common spike timing in subsets of channels can be used as a principle for grouping [Fig. 4(B)]. The lines remain labeled to signify the features they encode, while the time domain is used to signify which channels are grouped or separated [23], [24]. The temporal label can involve either timing relative to a reference wave [92]–[94] or spike synchronization between channels [19]–[22], [95].

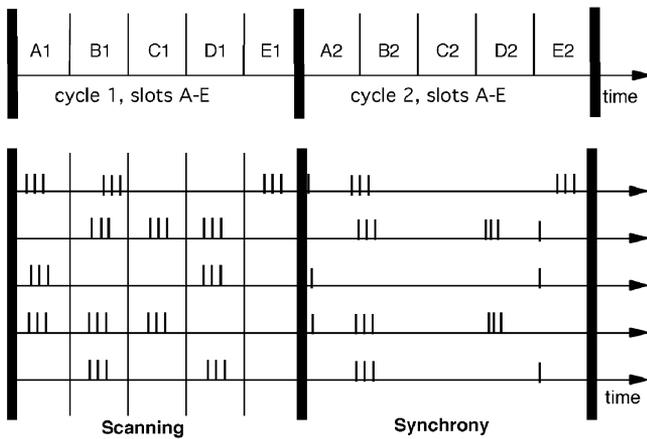
In frequency-division multiplexing [Fig. 4(B)], different signals utilize different frequency bands such that they can be mixed together in transmission channels and separated by receivers on the basis of their respective frequencies. In a pulse code, interpulse intervals encode different frequencies. The population-interval representation of low frequency sounds discussed above is a frequency-multiplexed system in that interspike intervals associated with different stimulus frequency components are conveyed by the same overlapping sets of auditory nerve fibers.

In code division multiplexing [Fig. 4(C)], patterns encoding features are not limited to any one frequency band. In the scheme shown in the figure, an initial header segment indicates signal-identity while a second segment conveys signal value. In a pulse code implementation, headers and value segments can be signaled by characteristic bursts consisting of different intra-burst intervals. The scheme here is similar to one proposed for multimodal encoding of cutaneous qualities [18]. Receivers selectively tuned to particular headers would process pulse trains segments with appropriate burst structure and ignore others. Multiple types of signals might then be asynchronously sent over each transmission line and demultiplexed by appropriate receivers.

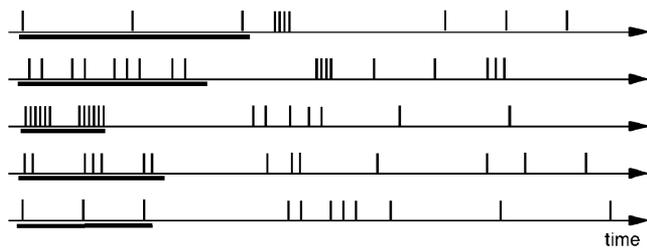
VII. NEURAL INFORMATION PROCESSING IN THE TIME DOMAIN

The near-ubiquity and robustness of spike timing information in early sensory representations begs questions concerning what neural temporal processing architectures are needed to fully exploit this information. The basic division between channel-based and temporal codes creates three broad classes of processing networks: connectionist networks, time-delay networks, and timing nets. Historically, time-delay networks have been used to convert analog, temporally-coded inputs into discrete patterns of channel-activations (e.g., [56], [84], [96]). However, neural networks can also be envisioned that carry out analog filtering and signal separation operations of sorts

A. Time-division multiplexing



B. Frequency-division multiplexing



C. Code-division multiplexing

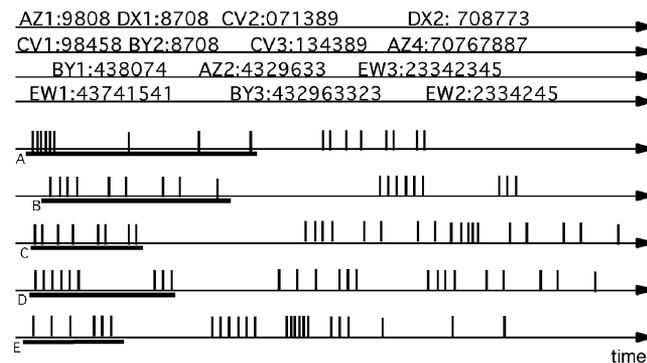


Fig. 4. Signal multiplexing: concurrent transmission of multiple signals over the same channels. (A) Time-division multiplexing (TDM) allocates particular time-slots to each signal (A-E). Neural pulse-coded TDM could use successive temporal windows or spike synchronies (B) Pulse-coded frequency-division multiplexing (FDM) maps signals onto different interspike intervals (bars). Multiple interval-types can then be transmitted over the same lines. (C) Code-division multiplexing encodes signal-type in the form of the signal, using header sequences that direct asynchronous and nonsequential transmission and reconstruction of the signal by each receiver. Code-division multiplexing using complex temporal pulse patterns as primitives. Initial burst encodes signal type, while subsequent spikes encode signal content. Patterns can be interleaved and transmitted asynchronously over the same channels.

familiar to signal processing engineers but relatively foreign to neuroscientists and neural network theorists. Using delay, multiplication (coincidence), and subtraction-cancellation [81] (inhibition) operations, neural timing networks carry out signal processing operations entirely in the time domain to compare, extract, and separate temporal patterns.

Feedforward timing nets (FFTNs) consist of arrays of idealized coincidence detectors that take inputs from multiple sets of tapped delay lines Fig. 5, [97]. Timing nets exploit the ability of axons to shift (delay) temporal patterns in time. Although structurally reminiscent of Jeffress and [56] and Braitenberg architectures [96], FFTNs have no subsequent “counting” or rate integration stage; their outputs are time rather than channel patterns. The computational properties of simple FFTNs with two sets of tapped delay input lines are straightforward. Spikes are fed in from either end of the array and propagate through their respective delay lines, crossing at different positions to produce output spikes. Each horizontal array position implements a particular relative delay between the input signals [Fig. 5(B)].

Several basic operations can be carried out [97]. First, the crosscorrelation function of the two inputs can be computed by counting the number of spikes in each output channel (vertical bar) as a function of relative delay (i.e., the Jeffress model). Convolutions can be computed by summing across relative delay channels for each time step [98], i.e., forming the collective peristimulus time histogram (PSTH) of the detector ensemble. The operation is similar to the common flip/shift/multiply method of computing convolution.

The “summary autocorrelation” of the outputs is the sum of the output channel autocorrelations (all-order interval distributions), analogous to the neural population-interval representations for pitch and timbre discussed above. Since the output summary autocorrelation of the FFTN is the product of the input autocorrelations, the coincidence array computes a cross spectral product in the time domain, from interval (lag) statistics. This multiplication of autocorrelations permits extraction of intervals related to common fundamentals (pitch) irrespective of the particular partials involved (timbre, vowel quality), and vice versa. Intervals associated with a common F0 shared by two signals can be extracted even if the signals have no harmonics in common.

A further consequence is that each interval or higher order spike arrival pattern (e.g., triplets) must be present in each of the inputs in order to appear in an output channel [Fig. 5(E)]. The array, thus, functions as a temporal sieve, passing those temporal patterns that are common to both sets of inputs, even when they are embedded in other spikes. These sifting operations are useful for computing intersections of sets of patterns, detecting embedded patterns, and for extracting patterns of interest from signals. Feedforward nets also separate temporal patterns on the basis of time-of-arrival disparities [Fig. 5(F)].

VIII. RECURRENT NEURAL TIMING NETS (RTNs)

Recurrent timing nets (RTNs) were developed to handle the detection and separation of periodic patterns. Periodic signals, such as those that produce strong pitches and rhythms, create strong temporal expectations. RTNs were inspired by time domain spike processing [98], stabilized auditory image architectures [99], reverberating circuits [100], neural loops [101], adaptive timing nets [102], Grossberg’s adaptive resonance circuits, and the psychology of temporal expectation. Concrete guidance has also come from auditory scene analysis: how perceptual attributes of sounds are organized into objects, voices, and streams

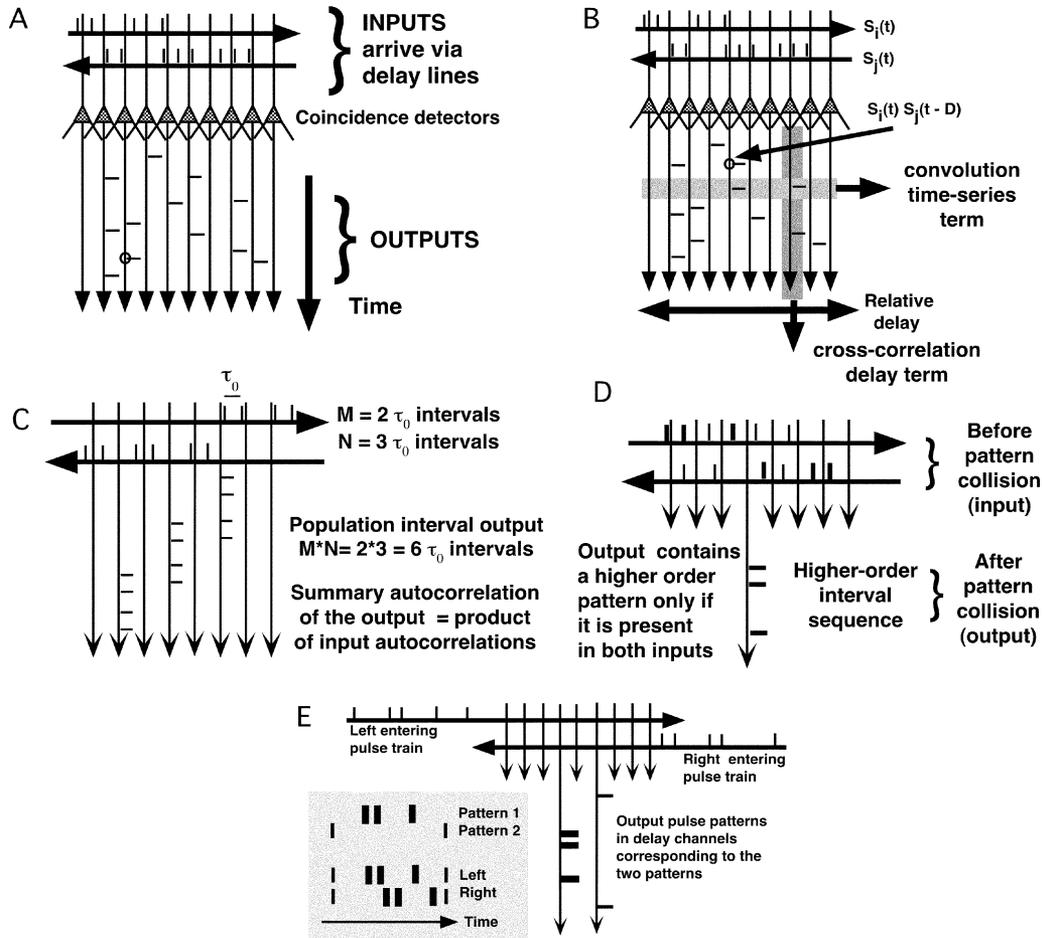


Fig. 5. Feedforward timing nets. (A) Basic structure. (B) Crosscorrelation by summing output pulses within channels at each relative delay. Convolution by summing across channels at each time step. (C) Summary autocorrelation of the output computes products of input autocorrelations. (D) Extraction of higher order temporal patterns containing interspersed spikes. (E) Separation of patterns (inset) with different relative delays or phases.

that can be separated from one another [103], [104]. Common periodicity (harmonicity) is a very strong grouping cue in audition. Although most low-frequency auditory perception is famously insensitive to the stationary phase-spectrum of the stimulus, transient phase and amplitude shifts are known to cause individual harmonics to separate from a complex. In the case of mistuned harmonics, which also readily separate, the phase of the mistuned component constantly precesses relative to that of the rest of the complex. In echolocation the neuronal mechanisms likely involve fine time comparisons between a signal (cry) and its repetition (echo) [60]. All of these examples suggest a process by which a signal waveform is compared with itself on a period-by-period basis.

In RTNs signals circulate in reverberating conduction loops, as temporal memory traces that interact with incoming temporal patterns. Neural representations then build up over time, with previous patterns dynamically creating sets of perceptual expectations that can either be confirmed and built up or violated. RTNs provide a basic mechanism for modeling echoic memory, separation of periodic sounds, rhythm induction, and creation of temporal expectancies.

Recurrent timing nets are in formative stages of development. The simplest recurrent timing networks consist of a 1-D array of coincidence detectors having common direct inputs [Fig. 6(A)].

The output of each coincidence element is fed into a recurrent delay line such that the output of the element at time t circulates through the line and arrives at the element at time $t + \tau$ [Fig. 6(B)]. In more recent implementations [105], a simple error-adjustment processing rule (1,2) governs the interaction of direct and circulating inputs

$$H(t) = H(t - \tau) + B_{\tau} [X(t) - H(t - \tau)] \quad (1)$$

$$B_{\tau} = \frac{\tau}{33 \text{ ms}}. \quad (2)$$

Here $X(t)$ is the direct input signal, $H(t - \tau)$ is the incoming circulating signal, and $H(t)$ is the outgoing circulating signal. τ is the recurrence time of the loop. B_{τ} determines the rate of adjustment and the length of the temporal processing window, which corresponds to the lower limit of pitch [81], [106].

In signal processing terms, RTNs most resemble neural time-domain implementations of comb filtering operations. Although the computations bear many similarities to correlogram-based segregation and oscillator-based channel allocations [95], by keeping processing entirely in the time domain, the RTN produces separated waveforms rather than ensembles of grouped frequency channels.

Some basic signal separation capabilities of RTNs have been tested on concurrently-presented vowel pairs (“double vowels”).

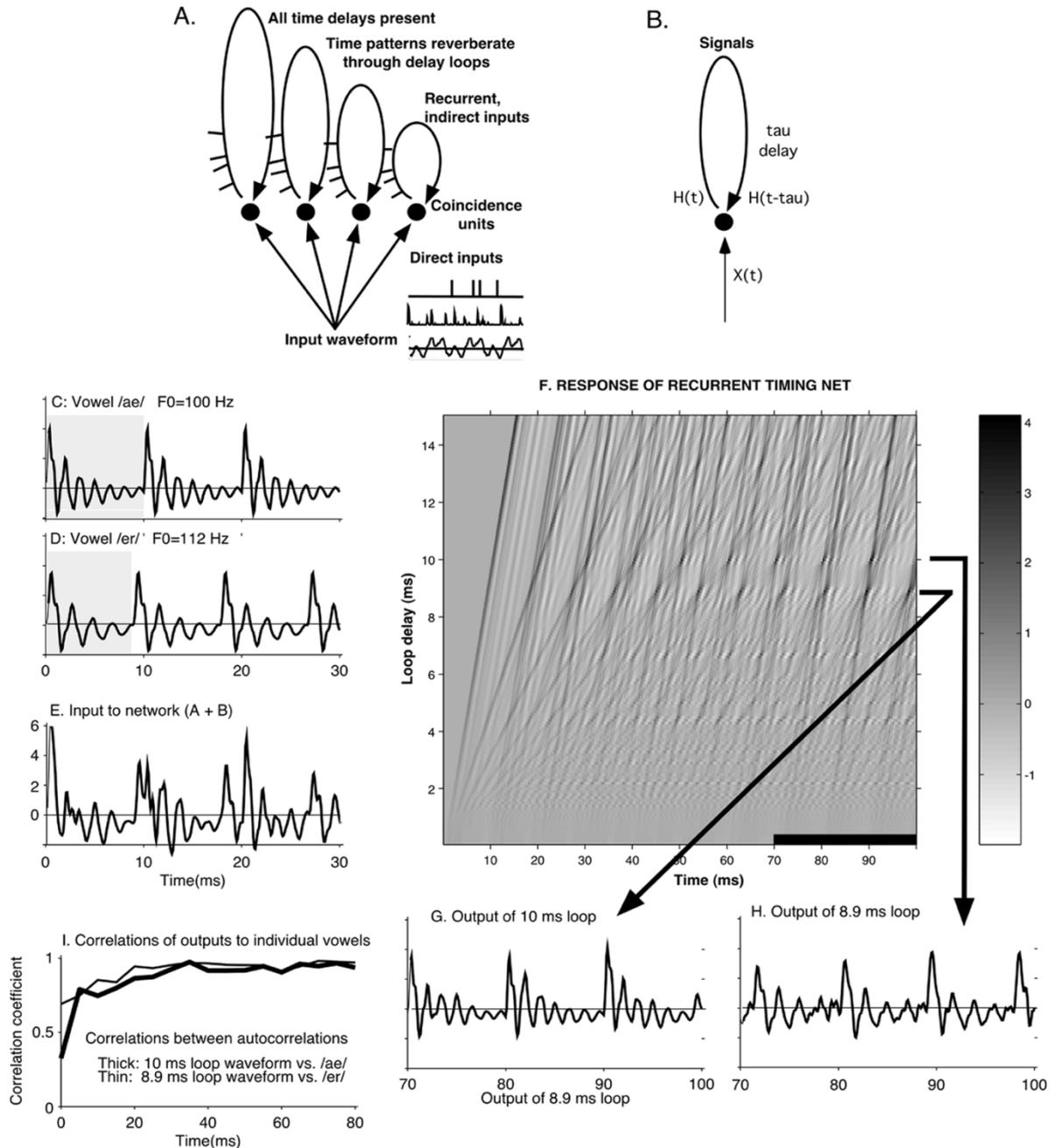


Fig. 6. Recurrent timing nets for F0-based separation of auditory objects. (A) Basic structure. (B) Processing nodes compare incoming and delayed (circulating) signals, and adjust output signal according to a linear error-adjustment rule. (C), (D) Waveforms of constituent single vowel waveforms. (E) Double vowel waveform. (F) Response of the network as a function of loop-delay and time. (G), (H) Circulating waveforms in the two most highly activated delay channels. (I) Resemblance of the autocorrelations (power spectra) of these two circulating waveforms to those of the individual constituent vowels.

A considerable body of psychophysical, neurophysiological, and neurocomputational work exists on their perception, [97]. When fundamentals are the same, the vowels are fused together, making the constituents harder to identify. When fundamentals are separated by a semitone (6%) or more, they are heard as two separate auditory objects, and can be identified with somewhat higher accuracy (15%–20% improvement).

A typical case is illustrated here (Fig. 6). Synthetic, three-formant double vowels (/ae/, /er/) with different fundamentals (100, 112 Hz, Fig. 8AB) were summed [Fig. 6(E)] and processed by a network with 150 delay loops that spanned 0–15 ms delays. Invariably, the two periodic patterns build up fastest in the delay loop whose recurrence time best matches their rep-

etition time [Fig. 6(F)]. Here the loops with the highest average signal strength correspond to the periods of the two vowels (8.9 and 10 ms). The signals circulating in these two delay channels after 70 ms of processing [Fig. 6(G) and (H)] highly resemble the two vowel constituents [Fig. 6(A) and (B)]. Correlations between the autocorrelations of separated signals and individual vowels [Fig. 6(I)] show how these resemblances increase as the signal separation unfolds over processing time. Multichannel RTNs with auditory nerve front-ends have also been explored, with similar results [97].

In effect, the loops dynamically create matched filters from repeating temporal patterns in the stimulus. Temporal-pattern invariances are then enhanced relative to uncorrelated patterns

in each loop, such that each loop functions as a pattern-amplifier for stimulus patterns whose period equals the loop's recurrence time. The network as a whole will amplify any periodic temporal patterns in its input and sort them out according to their periods. With longer sets of delays, such networks also separate multiple rhythm subpatterns [106].

RTNs illustrate how phase-locked fine timing information (and transient changes in phase) can be used to separate sounds. Most existing strategies for separating sounds on the basis of fundamental frequency attempt to group frequency channels together by finding F0-related features in each channel, e. g., [21], [22], [95], [107]. The RTN model demonstrates an alternate, time-domain strategy for auditory object separation that 1) exploits fine timing information and 2) uses no explicit feature detection (i.e., F0-detectors). Rather than labeling frequency channels for segregation, the delay loops amplify temporal pattern-invariances in the waveform. Auditory objects are, thus, separated on the basis of temporal pattern coherence [23]. Recurrent timing nets demonstrate how auditory images can be built up by comparing a signal with its immediate past [99]. They show how purely temporal representations and computations can effect separation and identification of auditory objects even when the information that constitutes them is temporally multiplexed in many of the same frequency channels. The nets provide a simple processing scheme in which object formation comes prior to analysis of auditory qualities (pitch, timbre) rather than as its consequence. They illustrate how auditory separation mechanisms might use transient changes in phase to form objects that are subsequently analyzed by phase-insensitive mechanisms for periodicity (pitch) and spectrum (timbre). In the context of speech separation and recognition, it may therefore make sense not to discard phase information in early stages of signal processing that may provide a basis for speaker separation.

Looking ahead, it is conceivable that judicious adjustment of synaptic weights and delays could result in networks capable of regenerating distributed temporal patterns—reverberating memories [100]. In randomly-connected nets, delay loops could be formed on the fly by short-term facilitations created by temporal correlations. The time structure of an incoming signal would dynamically organize central neural circuits so as to propagate and build up stable, reverberating patterns.

IX. POTENTIAL IMPLICATIONS FOR NEURAL NETWORKS

Temporal codes and computations may provide new avenues for artificial neural network development. Some of the general advantages of temporal codes stem from their ability to encode sensory information in a manner largely independent of stimulus intensity; this facilitates formation of intensity-invariant perceptual equivalence classes, which drastically simplifies pattern recognition. Signal multiplexing, mass statistical representations, and vectorial encodings may permit neural networks to handle information in more flexible ways that liberate signals from wires.

In a traditional connectionist network, all signals are scalars that are weighted and combined with many other signals at each node. The ability to temporally multiplex multiple independent

signals lessens their competitive interference. Networks can be envisioned in which neural assemblies tuned to respond to specific temporal patterns propagate both locally-relevant and irrelevant parts of neural signals. Inter-transparency of signals would then allow broadcast strategies of neural integration and coordination that would begin to resemble asynchronous, decentralized communications systems.

Temporal codes permit neural representations based on mass statistics of spike correlations. Population-interval representation of auditory qualities discussed above is a concrete example. If information can be encoded in temporal patterns of spikes, then processing can occur on population or ensemble-wide scales that obviate the need for particular elements, interconnections, synaptic efficacies, and transmission paths in order to function, a longstanding problem in theoretical neuroscience. Feedforward timing nets provide a simple processing strategy that obviates the need for precise point-to-point connectivities; although very primitive, they suggest strategies by which information might be processed on a mass-statistical basis without regard for which particular elements and connections are involved.

Finally, temporal codes support multiple, independent signal types. Independent signal types allow for vectorial, multidimensional representations in which temporal patterns related to different aspects of a situation or object might be flexibly superposed [20]. For example, characteristic temporal patterns might distinguish information from different sense modalities [18]. Tag systems based on orthogonal sets of temporal patterns might then permit representational compositionality and high-dimensional computations on them. Over the course of processing incoming information, neural assemblies would add specific annotational tags to circulating signals, such that the circulating temporal patterns could increase their effective dimensionality over time to create new categories.

REFERENCES

- [1] D. Marr, *Vision*. San Francisco, CA: W. H. Freeman, 1982.
- [2] D. H. Perkel and T. H. Bullock, "Neural coding," *Neurosci. Res. Program Bull.*, vol. 6, pp. 221–348, 1968.
- [3] W. R. Uttal, *The Psychobiology of Sensory Coding*. New York: Harper & Row, 1973.
- [4] D. M. MacKay and W. S. McCulloch, "The limiting information capacity of a neuronal link," *Bull. Math. Biophys.*, vol. 14, 1952.
- [5] A. Brand, O. Behrend, T. Marquardt, D. McAlpine, and B. Grothe, "Precise inhibition is essential for microsecond interaural time difference coding," *Nature*, vol. 417, pp. 543–7, 2002.
- [6] W. R. Uttal, *Sensory Coding: Selected Readings*. Boston, MA: Little-Brown, 1972, pp. 481–481.
- [7] F. Theunissen and J. P. Miller, "Temporal encoding in nervous systems: A rigorous definition," *J. Comput. Neurosci.*, vol. 2, pp. 149–62, 1995.
- [8] J. D. Victor and K. Purpura, "Metric-space analysis of spike trains: Theory, algorithms, and application," *Network*, vol. 8, pp. 127–164, 1997.
- [9] J. J. Eggermont, *The Correlative Brain: Theory and Experiment in Neural Interaction*. Berlin, Germany: Springer-Verlag, 1990, vol. 24.
- [10] W. Gerstner, "Spiking neurons," in *Pulsed Neural Networks*, W. Maas and C. M. Bishop, Eds. Cambridge, MA: MIT Press, 1999, pp. XIII–XXVI.
- [11] W. Maas, "Computing with spiking neurons," in *Pulsed Neural Networks*, W. Maas and C. M. Bishop, Eds, MA, Cambridge: MIT Press, 1999, pp. XIII–XXVI.
- [12] P. Cariani, "As if time really mattered: Temporal strategies for neural coding of sensory information," *Commun. Cogn. Artif. Intell.*, vol. 12, pp. 161–229, 1995.

- [13] ———, "Temporal coding of sensory information in the brain," *Acoust. Sci. Tech.*, vol. 22, pp. 77–84, 2001.
- [14] B. J. Richmond, L. M. Optican, and T. J. Gawne, "Neurons use multiple messages encoded intemporally modulated spike trains to represent pictures," in *Seeing Contour and Color*, J. J. Kulikowski and C. M. Dickenson, Eds. New York: Pergamon, 1989, pp. 705–713.
- [15] J. D. Victor and K. Purpura, "Nature and precision of temporal coding in visual cortex," *J. Neurophysiol.*, vol. 76, pp. 1310–1326, 1996.
- [16] B. J. Richmond and T. J. Gawne, "The relationship between neuronal codes and cortical organization," in *Neuronal Ensembles: Strategies for Recording and Decoding*, H. B. Eichenbaum and J. L. Davis, Eds. New York: Wiley, 1998, pp. 57–80.
- [17] V. Mountcastle, "The problem of sensing and the neural coding of sensory events," in *The Neurosciences: A Study Program*, G. C. Quarten, T. Melnechuk, and F. O. Schmitt, Eds. New York: Rockefeller Univ. Press, 1967.
- [18] R. Emmers, *Pain: A Spike-Interval Coded Message in the Brain*. New York: Raven, 1981.
- [19] C. M. Gray, "The temporal correlation hypothesis of visual feature integration: Still alive and well," *Neuron*, vol. 24, pp. 31–47, 1999.
- [20] W. Singer, "Neuronal synchrony: A versatile code for the definition of relations?," *Neuron*, vol. 24, pp. 49–65, 1999.
- [21] D. Wang, "The time dimension for neural computation," Dept. Comput. Inform. Sci., The Ohio State Univ., Columbus, OH, OSU-CISRC-09/02/TR21, 2002.
- [22] D. L. Wang, "An oscillatory correlation theory of temporal pattern segmentation," in *Neural Representation of Temporal Patterns*, E. Covey, H. L. Hawkins, and R. F. Port, Eds. New York: Plenum, 1995, pp. 53–76.
- [23] C. von der Malsburg, "The correlation theory of brain function," in *Models of Neural Networks II. Temporal Aspects of Coding and Information Processing in Biological Systems*, E. Doumany, J. L. van Hemmen, and K. Schulten, Eds. New York: Springer-Verlag, 1994, pp. 95–119.
- [24] L. Shastri and V. Ajanagadde, "From simple associations to systematic reasoning: A connectionist representation of rules, variables, and dynamic bindings using temporal synchrony," *Behav. Brain Sci.*, vol. 16, pp. 417–494, 1993.
- [25] G. S. Wasserman, "Isomorphism, task dependence, and the multiple meaning theory of neural coding," *Biol. Signals*, vol. 1, pp. 117–142, 1992.
- [26] S. H. Chung, S. A. Raymond, and J. Y. Lettvin, "Multiple meaning in single visual units," *Brain Behav. Evol.*, vol. 3, pp. 72–101, 1970.
- [27] R. Lestienne and B. L. Strehler, "Time structure and stimulus dependence of precise replicating patterns present in monkey cortical neuronal spike trains," *Brain Res.*, vol. 43, pp. 214–238, 1987.
- [28] M. Abeles, H. Bergman, E. Margalit, and E. Vaadia, "Spatiotemporal firing patterns in the frontal cortex of behaving monkeys," *J. Neurophysiol.*, vol. 70, pp. 1629–1638, 1993.
- [29] S. S. Stevens, "Sensory power functions and neural events," in *Principles of Receptor Physiology*, W. R. Loewenstein, Ed. Berlin, Germany: Springer-Verlag, 1971, pp. 226–242.
- [30] P. Heil, "Auditory cortical onset responses revisited. I. First-spike timing," *J. Neurophysiol.*, vol. 77, pp. 2616–2641, 1997.
- [31] J. P. Segundo, G. P. Moore, L. J. Stensaas, and T. H. Bullock, "Sensitivity of neurones in Aplysia to temporal pattern of arriving impulses," *J. Exp. Biol.*, vol. 40, pp. 643–667, 1963.
- [32] W. R. Uttal, *The Psychobiology of Mind*. Hillsdale, NJ: Lawrence Erlbaum, 1978.
- [33] G. Werner and V. B. Mountcastle, "The variability of central neural activity in a sensory system, and its implications for the central reflection of sensory events," *J. Neurophysiol.*, vol. 26, pp. 958–977, 1963.
- [34] L. Abbott and T. J. Sejnowski, "Introduction," in *Neural Codes and Distributed Representations: Foundations of Neural Computation*, L. Abbott and T. J. Sejnowski, Eds. Cambridge, MA: MIT Press, 1999, pp. 225–242.
- [35] S. A. Deadwyler and R. E. Hampson, "Ensemble activity and behavior: What's the code?," *Science*, vol. 270, pp. 1316–1318, 1995.
- [36] P. A. Cariani and B. Delgutte, "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. Neurophysiol.*, vol. 76, pp. 1698–1734, 1996.
- [37] A. Przybyszewski, J. P. Gasa, W. Foote, and D. A. Pollen, "Striate cortex increases contrast gain in macaque LGN neurons," *Vis. Neurosci.*, vol. 17, pp. 485–494, 2000.
- [38] W. Keidel, "The sensory detection of vibrations," in *Foundations of Sensory Science*, W. W. Dawson and J. M. Enoch, Eds. Berlin, Germany: Springer-Verlag, 1984, pp. 465–512.
- [39] V. Mountcastle, "Temporal order determinants in a somatosensory frequency discrimination: Sequential order coding," *Ann. New York Acad. Sci.*, vol. 682, pp. 151–170, 1993.
- [40] F. Macrides and S. L. Chorover, "Olfactory bulb units: Activity correlated with inhalation cycles and odor quality," *Science*, vol. 175, pp. 84–86, 1972.
- [41] C. E. Carr, "Processing of temporal information in the brain," *Annu. Rev. Neurosci.*, vol. 16, pp. 223–243, 1993.
- [42] M. A. Frye, "Encoding properties of the wing hinge stretch receptor in the hawkmoth *manduca sexta*," *J. Exp. Biol.*, vol. 204, pp. 3693–702, 2001.
- [43] G. Laurent, "A systems perspective on early olfactory coding," *Science*, vol. 286, pp. 723–8, 1999.
- [44] J. S. Kauer, "Response patterns of amphibian olfactory bulb neurones to odour stimulation," *J. Physiol.*, vol. 243, pp. 695–715, 1974.
- [45] G. Laurent and H. Davidowitz, "Encoding of olfactory information with oscillating neural assemblies," *Science*, vol. 265, pp. 1872–1875, 1994.
- [46] J. White, T. A. Dickinson, D. R. Walt, and J. S. Kauer, "An olfactory neuronal network for vapor recognition in an artificial nose," *Biol. Cybern.*, vol. 78, pp. 245–251, 1998.
- [47] E. Covey, *Temporal Neural Coding in Gustation*. Durham, NC: Duke Univ. Press, 1980.
- [48] P. M. Di Lorenzo and G. S. Hecht, "Perceptual consequences of electrical stimulation in the gustatory system," *Behav. Neurosci.*, vol. 107, pp. 130–138, 1993.
- [49] B. J. Richmond, L. M. Optican, M. Podell, and H. Spitzer, "Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. I. Response characteristics," *J. Neurophysiol.*, vol. 57, pp. 132–146, 1987.
- [50] W. M. Kozak and H. J. Reitboeck, "Color-dependent distribution of spikes in single optic tract fibers of the cat," *Vis. Res.*, vol. 14, pp. 405–419, 1974.
- [51] J. J. Sheppard, *Human Color Perception: A Critical Study of the Experimental Foundation*. New York: Elsevier, 1968.
- [52] L. Festinger, M. R. Allyn, and C. W. White, "The perception of color with achromatic stimulation," *Vis. Res.*, vol. 11, pp. 591–612, 1971.
- [53] R. A. Young, "Some observations on temporal coding of color vision: Psychophysical results," *Vis. Res.*, vol. 17, pp. 957–965, 1977.
- [54] G. von Békésy, *Sensory Inhibition*. Princeton, NJ: Princeton Univ. Press, 1967.
- [55] T. H. Bullock, "Signals and neural coding," in *The Neurosciences: A Study Program*, G. C. Quarten, T. Melnechuk, and F. O. Schmitt, Eds. New York: Rockefeller Univ. Press, 1967, pp. 347–352.
- [56] L. A. Jeffress, "A place theory of sound localization," *J. Comp. Physiol. Psychol.*, vol. 41, pp. 35–39, 1948.
- [57] S. Colburn, "Computational models of binaural processing," in *Auditory Computation*, H. Hawkins, T. McMullin, A. N. Popper, and R. R. Fay, Eds. New York: Springer-Verlag, 1996.
- [58] M. Konishi, "Deciphering the brain's codes," in *Neural Codes and Distributed Representations: Foundations of Neural Computation*, L. Abbott and T. J. Sejnowski, Eds. Cambridge, MA: MIT Press, 1999, pp. 1–18.
- [59] C. F. Moss and H.-U. Schnitzler, "Behavioral studies of auditory information processing," in *Hearing by Bats*, A. N. Popper and R. R. Fay, Eds. New York: Springer-Verlag, 1995, pp. 87–145.
- [60] J. A. Simmons, "Formation of perceptual objects from the timing of neural responses: Target-range images in bat sonar," in *The Mind-Brain Continuum*, R. Llinas and P. S. Churchland, Eds. Cambridge, MA: MIT Press, 1996, pp. 219–250.
- [61] W. Heiligenberg, "The coding and processing of temporal information in the electrosensory system of the fish," in *Temporal Coding in the Brain*, G. Buzaki, R. Llinas, W. Singer, A. Berthoz, and Y. Christen, Eds. Berlin, Germany: Springer-Verlag, 1994, pp. 1–12.
- [62] W. Reichardt, "Autocorrelation, a principle for the evaluation of sensory information by the central nervous system," in *Sensory Communication*, W. A. Rosenblith, Ed. New York: Wiley, 1961, pp. 303–317.
- [63] W. Bialek, F. Rieke, R. R. van Steveninck, and W. D. de Ruyter, "Reading a neural code," *Science*, vol. 252, pp. 1854–1856, 1991.
- [64] F. Rieke, D. Warland, R. de Ruyter van Steveninck, and W. Bialek, *Spikes: Exploring the Neural Code*. Cambridge, MA: MIT Press, 1997.
- [65] D. A. Pollen, J. P. Gaska, and L. D. Jacobson, "Physiological constraints on models of visual cortical function," in *Models of Brain Function*, R. M. J. Cotterill, Ed. Cambridge, U.K.: Cambridge Univ. Press, 1989, pp. 115–136.
- [66] G. T. Plant, "Temporal properties of normal and abnormal spatial vision," in *Spatial Vision*, D. Regan, Ed. Boca Raton, FL: CRC Press, 1991, pp. 43–63.

- [67] S. P. Strong, R. Koberle, R. R. de Ruyter van Steveninck, and W. Bialek, "Entropy and information in neural spike trains," *Phys. Rev. Lett.*, vol. 80, pp. 197–200, 1998.
- [68] P. Reinagel and C. Reid, "Temporal coding of visual information in the thalamus," *J. Neurosci.*, vol. 20, pp. 5392–5400, 2000.
- [69] B. Buracas, A. M. Zador, M. R. DeWeese, and T. D. Albright, "Efficient discrimination of temporal patterns by motion-sensitive neurons in primate visual cortex," *Neuron*, vol. 20, pp. 959–969, 1998.
- [70] A. Kreiter and W. Singer, "Stimulus-dependent synchronization of neuronal responses in the visual cortex of the awake macaque monkey," *J. Neurosci.*, vol. 16, pp. 2381–2396, 1996.
- [71] R. W. Ditchburn and B. L. Ginsborg, "Vision with a stabilized retinal image," *Nature*, vol. 170, pp. 178–94, 1952.
- [72] F. Mechler, J. D. Victor, K. P. Purpura, and R. Shapley, "Robust temporal coding of contrast by V1 neurons for transient but not for steady-state stimuli," *J. Neurosci.*, vol. 18, pp. 6583–6598, 1998.
- [73] S.-H. Lee and R. Blake, "Visual form created solely from temporal structure," *Science*, vol. 284, pp. 1165–1168, 1999.
- [74] T. Carney, D. A. Silverstein, and S. A. Klein, "Vernier acuity during image rotation and translation: Visual performance limits," *Vis. Res.*, vol. 35, pp. 1951–1964, 1995.
- [75] P. Reinagel and R. C. Reid, "Precise firing events are conserved across neurons," *J. Neurosci.*, vol. 22, pp. 6837–6841, 2002.
- [76] W. R. Uttal, *On Seeing Forms*. Hillsdale, NJ: Lawrence Erlbaum, 1988.
- [77] —, *An Autocorrelation Theory of Form Detection*. New York: Wiley, 1975.
- [78] H. J. Reitboeck, M. Pabst, and R. Eckhorn, "Texture description in the time domain," in *Computer Simulation in Brain Science*, R. M. J. Cotterill, Ed. Cambridge, U.K.: Cambridge Univ. Press, 1988.
- [79] B. C. J. Moore, *An Introduction to the Psychology of Hearing*, Fifth ed. San Diego, CA: Academic, 2003.
- [80] A. de Cheveigne, "Time-domain auditory processing of speech," *J. Phonetics*, vol. 31, pp. 547–561, 2003.
- [81] A. de Cheveigné, "Pitch perception models," in *Pitch*, C. J. Plack and A. J. Oxenham, Eds. New York: Springer-Verlag.
- [82] J. L. Goldstein and P. Srulovicz, "Auditory-nerve spike intervals as an adequate basis for aural frequency measurement," in *Psychophysics and Physiology of Hearing*, E. F. Evans and J. P. Wilson, Eds. San Diego, CA: Academic, 1977.
- [83] B. Delgutte, "Physiological models for basic auditory percepts," in *Auditory Computation*, H. Hawkins, T. McMullin, A. N. Popper, and R. R. Fay, Eds. New York: Springer-Verlag, 1996.
- [84] J. C. R. Licklider, "A duplex theory of pitch perception," *Experientia*, vol. VII, pp. 128–134, 1951.
- [85] R. Lyon and S. Shamma, "Auditory representations of timbre and pitch," in *Auditory Computation*, H. Hawkins, T. McMullin, A. N. Popper, and R. R. Fay, Eds. New York: Springer-Verlag, 1995, pp. 517–517.
- [86] R. Meddis and L. O'Mard, "A unitary model of pitch perception," *J. Acoust. Soc. Am.*, vol. 102, pp. 1811–1820, 1997.
- [87] P. Cariani, "Temporal coding of periodicity pitch in the auditory system: An overview," *Neural Plasticity*, vol. 6, pp. 147–172, 1999.
- [88] A. R. Palmer, "Segregation of the responses to paired vowels in the auditory nerve of the guinea pig using autocorrelation," in *The Auditory Processing of Speech*, S. M. E. H., Ed. Berlin: Springer-Verlag, 1992, pp. 115–124.
- [89] T. Hirahara, P. Cariani, and B. Delgutte, "Representation of low-frequency vowel formants in the auditory nerve," in *Proc. ESCA Research Workshop Auditory Basis of Speech Perception*, U.K., July 15–19, 1996.
- [90] O. Ghitza, "Temporal nonplace information in the auditory-nerve firing patterns as a front-end for speech recognition in a noisy environment," *J. Phonetics*, vol. 16, pp. 109–123, 1988.
- [91] J. W. Morley, J. S. Archer, D. G. Ferrington, M. J. Rowe, and A. B. Turman, "Neural coding of complex tactile vibration," in *Information Processing in Mammalian Auditory and Tactile Systems*: Alan R. Liss, 1990, pp. 127–140.
- [92] J. E. Lisman and M. A. Idiart, "Storage of 7 ± 2 short-term memories in oscillatory subcycles," in *Science*, 1995, vol. 267, pp. 1512–5.
- [93] G. Laurent, M. Wehr, and H. Davidowitz, "Temporal representations of odors in an olfactory network," *J. Neurosci.*, vol. 16, pp. 3837–3847, 1996.
- [94] R. Llinás, U. Ribary, M. Joliot, and X.-J. Wang, "Context and content in temporal thalamocortical binding," in *Temporal Coding in the Brain*, G. Buzsáki, R. Llinás, W. Singer, A. Berthoz, and Y. Christen, Eds. Berlin, Germany: Springer-Verlag, 1994, pp. 251–272.
- [95] D. L. Wang and G. J. Brown, "Separation of speech from interfering sounds based on oscillatory correlation," *IEEE Trans. Neural Networks*, vol. 10, pp. 684–697, July 1999.
- [96] V. Braitenberg, "Functional interpretation of cerebellar histology," *Nature*, vol. 190, pp. 539–540, 1961.
- [97] P. Cariani, "Neural timing nets," *Neural Netw.*, vol. 14, pp. 737–753, 2001.
- [98] H. C. Longuet-Higgins, "A mechanism for the storage of temporal correlations," in *The Computing Neuron*, R. Durbin, C. Miall, and G. Mitchison, Eds. Reading, MA: Addison-Wesley, 1989, pp. 99–104.
- [99] R. D. Patterson, M. H. Allerhand, and C. Giguere, "Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform," *J. Acoust. Soc. Amer.*, vol. 98, pp. 1890–1894, 1995.
- [100] J. Orbach, *The Neuropsychological Theories of Lashley and Hebb*: Univ. Press America, 1998.
- [101] R. W. Thatcher and E. R. John, *Functional Neuroscience Vol. I. Foundations of Cognitive Processes*. Hillsdale, NJ: Lawrence Erlbaum, 1977.
- [102] D. M. MacKay, "Self-organization in the time domain," in *Self-Organizing Systems 1962*, M. C. Yovitts, G. T. Jacobi, and G. D. Goldstein, Eds. Washington, D.C.: Spartan, 1962, pp. 37–48.
- [103] A. S. Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, MA: MIT Press, 1990.
- [104] C. J. Darwin and R. P. Carlyon, "Auditory grouping," in *Handbook of Perception and Cognition: Hearing*, B. C. J. Moore, Ed. New York: Academic, 1995, pp. 387–424.
- [105] P. Cariani, "Recurrent timing nets for auditory scene analysis," in *Proc. Int. Joint Conf. Neural Networks (IJCNN)*, Portland, OR, July 20–24, 2003, [Online] <http://cil.ece.uic.edu/ijcnn03/>.
- [106] —, "Temporal codes, timing nets, and music perception," *J. New Music Res.*, vol. 30, pp. 107–136, 2002.
- [107] R. Meddis and M. J. Hewitt, "Modeling the perception of concurrent vowels with different fundamental frequencies," *J. Acoust. Soc. Amer.*, vol. 91, pp. 233–245, 1992.



Peter A. Cariani received the B.S. degree in life sciences from the Massachusetts Institute of Technology, Cambridge, in 1978 and the Ph.D. degree from Binghamton University, Binghamton, NY, in 1989. His thesis considered the epistemological implications of sensor evolution for artificial adaptive devices.

He is currently an Assistant Professor in Otolaryngology and Laryngology at Harvard Medical School, Cambridge, MA.

Chapter 4

Recurrent Timing Nets for F0-based Speaker Separation

Peter Cariani

Eaton Peabody Laboratory of Auditory Physiology

Massachusetts Eye & Ear Infirmary

243 Charles St., Boston, MA 02114 USA

cariani@mac.com

1 INTRODUCTION

Arguably, the most important barrier to widespread use of automatic speech recognition systems in real-life situations is their present inability to separate speech of individual speakers from other sound sources: other speakers, acoustic clutter, background noise. We believe that critical examination of biological auditory systems with a focus on "reverse-engineering" these systems, can lead to discovery of new functional principles of information representation and processing that can subsequently be applied to the design of artificial speech recognition systems.

From our experiences in attempting to understand the essentials of how the auditory system works as an information processing device, we believe that there are three major areas where speech recognizers could profit from incorporating processing strategies inspired by auditory systems. These areas are: 1) use of temporally-coded, front-end representations that are precise, robust, and transparent, and which encode the fine temporal (phase) structure of periodicities below 4 kHz; 2) use of early scene analysis mechanisms that form distinct auditory objects by means of common onset/offset/temporal contiguity and common harmonic structure (F0, voice pitch); and 3) use of central phonetic analyzers that are designed to operate on multiscale, temporally-coded, autocorrelation-like front-end representations as they present themselves after initial object formation/scene analysis processing. This paper will address the first two areas, with emphasis on possible neural mechanisms (neural timing nets) that could exploit phase-locked fine timing information to separate harmonic sounds on the basis of differences in their fundamental frequencies (harmonicity).

Psychoacoustical evidence suggests that the auditory system employs extremely effective low-level, bottom-up representational and scene analysis strategies to enable individual sound sources to perform this separation. Neurophysiological evidence suggests that the auditory system utilizes interspike interval information for representing sound in early stages of auditory processing. Interval-based temporal codes are known to provide high-quality, precise, and robust representations of stimulus periodicities and spectra over large dynamic ranges and in adverse sonic environments.

We have recently proposed *neural timing networks* that operate on temporally-coded inputs to carry out spike pattern analyses entirely in the time domain. These complement connectionist and time-delay architectures that

produce "spatial", atemporal patterns of element activations as their outputs. In effect neural timing architectures provide neural network implementations of analog signal processing operations (e.g. cross-correlation, autocorrelation, convolution, cross-spectral product). The ubiquity of neural (tapped) delay lines in the brain may mean that many signal processing operations are more easily and flexibly implemented neurally using time domain rather than frequency domain and/or discrete feature detection strategies.

We have found that simple *recurrent timing nets* can be devised that operate on temporal fine structure of inputs to build up and separate periodic signals with different fundamental periods (Neural Networks, 14: 737-753, 2001). Simple recurrent nets consist of arrays of coincidence detectors fed by common input lines and conduction delay loops of different recurrence times. A processing rule facilitates correlations between input and loop signals to amplify periodic patterns and segregate those with different periods, thereby allowing constituent waveforms to be recovered. The processing is akin to a dense array of comb filters with short temporal windows. Based on time codes and temporal processing, timing nets constitute a new, general strategy for scene analysis in neural networks. The nets build up correlational invariances rather than using features to label, segregate and bind channels. They provide a possible means by which the fine temporal structure of voiced speech might be exploited for the speaker separation and enhancement.

2 PITCH AND AUDITORY SCENE ANALYSIS

Perhaps the most basic function of a perceptual system is to coherently organize the incoming flux of sensory information into separate stable objects (Bregman, 1981; Bregman, 1990; Handel, 1989; Mellinger and Mont-Reynaud, 1996). In hearing, sound components are fused into unified objects, streams and voices that exhibit perceptual attributes, such as pitch, timbre, loudness, and location. Common periodicity, temporal proximity (onset, duration, offset), frequency, amplitude dynamics, phase coherence, and location in auditory space are some of the factors that contribute to fusions and separations of sounds.

For concurrent sounds, common harmonic structure plays perhaps the strongest role in forming unified objects and separating them (Hartmann, 1988; Mellinger and Mont-Reynaud, 1996). As a rule of thumb, voices and musical instruments having different fundamental frequencies (F0s) can be easily separated provided that their fundamentals differ by more than a semitone (6%, the F0 separation of two adjacent keys on a piano). Harmonic complexes with different fundamentals produce strong pitches at their fundamentals (even when the fundamental is not present in the power spectrum or when frequencies near the fundamental are masked out with noise). The mechanisms underlying pitch perception and auditory object formation therefore appear to intimately linked.

3 PROBLEMS WITH TRADITIONAL CHANNEL-BASED CODING OF THE AUDITORY STIMULUS

Traditionally, following the formidable intellectual syntheses of Helmholtz and Fletcher, the auditory system has been conceived in terms of a central analysis of

running frequency-domain representations of the stimulus power spectrum. Almost all extant front-end representations for automatic speech recognizers follow this assumption that a magnitude time-frequency spectrograph-like representation of the speech signal crudely mirrors the processing taking place in the central auditory system. Although it is often conventionally assumed that the cochlea itself implements an array of narrowly-tuned bandpass filters that then subsequently create tonotopically-organized frequency maps in the central auditory system, this picture of spectral coding presents many profound difficulties in light of what we know about neural response patterns at the level of the auditory nerve and the auditory CNS.

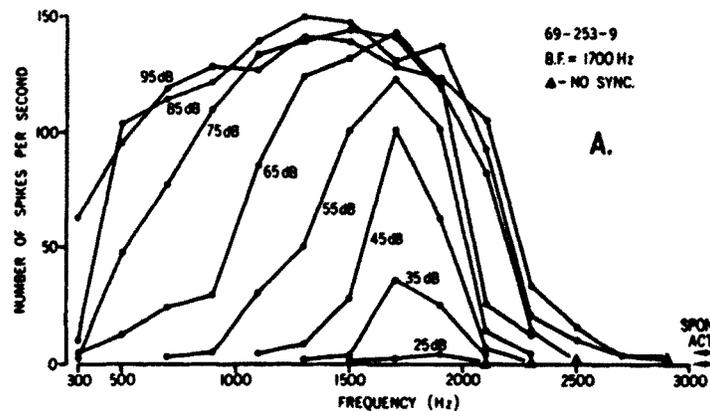


Fig. 4.1. Discharge rate of a low spontaneous rate, high threshold auditory nerve fiber as a function of tone frequency and sound pressure level. From (Rose et al., 1971), used with permission.

First, it has been known for decades now (Fig. 4.1) that the frequency response areas of auditory nerve fibers broaden dramatically as stimulus levels exceed 60 dB SPL (where tones below a fiber's characteristic frequency lie in the broad "tail" region of the fiber's tuning curve). A direct consequence of this is that simple representations of the stimulus spectrum based upon profiles of firing rates across whole neural populations (Fig. 4.2) perform best at low sound pressure levels ($\ll 50$ dB SPL).

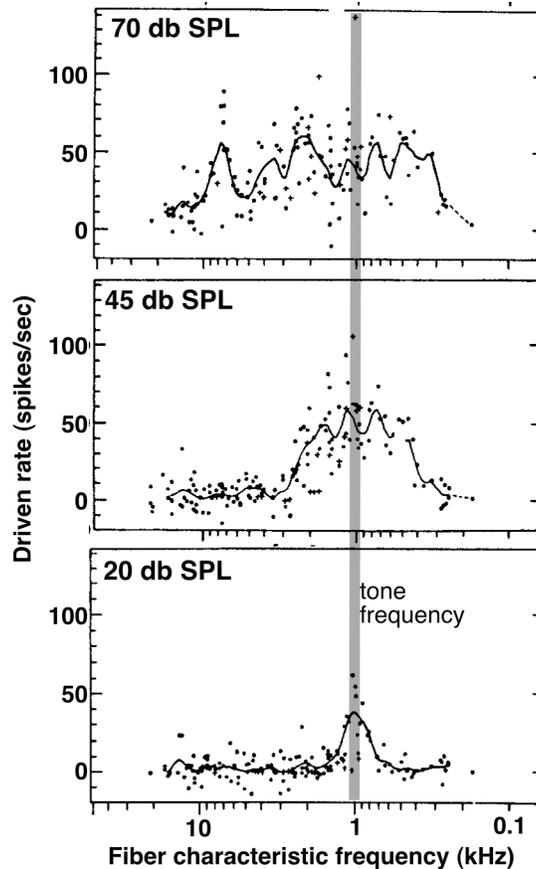


Figure 4.2. Rate-place profiles for a population of cat auditory nerve fibers of different characteristic frequencies in response to a 1 kHz tone presented at three sound pressure levels. Driven rate is average firing rate in response to a tone minus average firing rate in quiet. Plots from (Kim and Molnar, 1979), used by permission. See also (Irvine, 1986).

Such representations will be degraded at moderate to high SPLs as firing rates of fibers saturate. This is in marked contrast to auditory percepts, which almost invariably improve in precision and reliability at higher stimulus levels. In functional terms, this tonotopy is only viable as a scheme for encoding fine spectral distinctions at low sound levels. For frequencies most important for music and speech, tonotopy is therefore better conceptualized as "cochleotopy" – as a reflection of the most direct connections to the sensory receptor surface, rather than as a vehicle for frequency coding in its own right. Throughout the auditory neurophysiology literature, this general problem with the degradation of tonotopy, which is most acute for best frequencies below a few kHz, is seen at all levels of auditory processing from periphery to cortex.

Two general (arguably ad-hoc) remedies have been proposed to save the rate-place picture: 1) the selective use of the fewer and fewer neural elements that exist that have higher and higher rate-thresholds and therefore dynamic ranges better suited to encode higher stimulus levels, and 2) reliance on lateral inhibition and the locations of shoulders of population-wide excitation patterns to infer the

stimulus spectrum. Although these strategies can be made to work in particular cases (e.g. 2AFC discrimination), it is hard to envision how such representations would work in more general contexts to infer the spectra of novel stimuli and multiple sounds. It is even harder to imagine why pitch and timbre percepts based on such representations and analyses would be highly invariant with respect to stimulus intensity, with jnd's even improving at levels where units with appropriately high thresholds are relatively few.

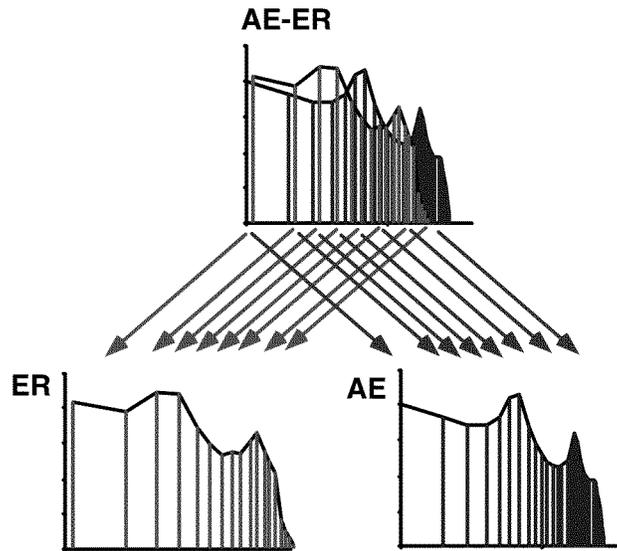


Figure 4.3. Traditional strategies for auditory scene analysis operate in the frequency domain to assign different sets of frequency channels to different auditory objects.

Traditional approaches to auditory scene analysis also operate using the central spectrum assumption, that auditory objects are best described in terms of patterns of activation amongst frequency channels. In this view, the task of an auditory separation mechanism is channel segregation and binding (Fig. 3), i.e. to label and segregate frequency channels according to pitch-related features (e.g. (Meddis and Hewitt, 1992)) and then to bind them together to form separate objects, streams, and voices.

4 TEMPORAL REPRESENTATION OF AUDITORY STIMULI

Historically, conceptions of the auditory system as a temporal pattern (periodicity) analyzer have developed alongside those that cast it as a frequency analyzer (Boring, 1942; Cariani, 1999; de Cheveigne, 2004). When one examines patterns of neural activity in the auditory nerve array (Fig. 4.4), one is immediately struck by the ubiquity of the temporal patterning of activity. In essence, the stimulus impresses its fine time structure on the temporal discharge patterns of multitudes of nerve fibers.

Periodicity-based theories of auditory representation account for the pitches of complex tones in terms of population-based all-order interspike interval statistics

(Cariani & Delgutte, 1996; Cariani, 1999). The most common sets of intervals present in the whole population at any given time predict the pitch that will be heard. This account encompasses pitches evoked by pure and complex tones that have periodicities (frequencies, fundamentals) below the limits of strong phase-locking (~ 4 kHz). The one exception of which we are aware are click rate pitches produced by click trains of alternating polarity, where the click rate pitch is heard an octave above the fundamental. Periodicity-based representations effectively explain the frequency limits of musical tonality (octave matching, musical interval recognition), which only exists up to around 4 kHz (Burns, 1999).

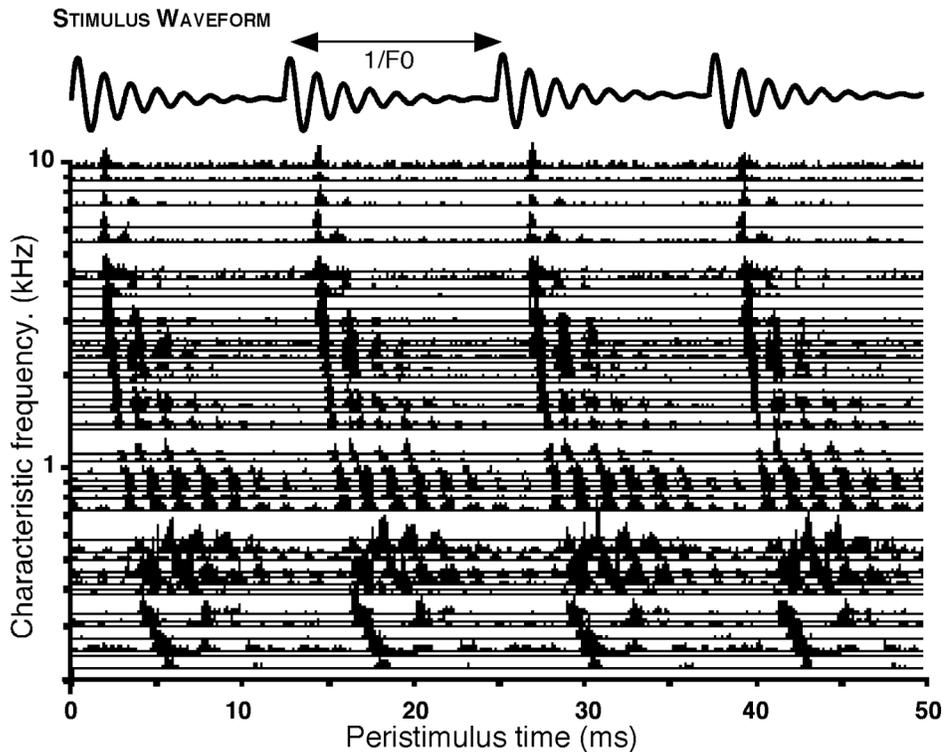


Figure 4.4. Temporal discharge patterns of auditory nerve fibers to a complex harmonic stimulus. Peristimulus time histogram (PSTH) responses of 52 cat auditory nerve fibers to a single-formant vowel (F_0 : 80 Hz, $F_1 = 640$ Hz, 100x, 60 dB SPL). PSTH baselines indicate fiber characteristic frequencies.

For example, the familiar Big Ben melody can be played in different descending registers, beginning around 10 kHz. At about 4 kHz, the melody becomes recognizable. The level-invariant nature of low-frequency hearing (vs. the level-dependent character of high frequency hearing) strongly suggests that low- and high-frequency hearing rely on different neural mechanisms: a temporal mechanism for periodicities below 4 kHz, and an atonal, level-dependent mechanism for higher frequencies.

Lower frequency spectra are also effectively represented in population-interval distributions. The reason is that each partial impresses its temporal structure on the temporal discharge patterns of fibers whose CFs are closest to it in frequency.

The result is that the stimulus partitions cochlear territories according to the relative

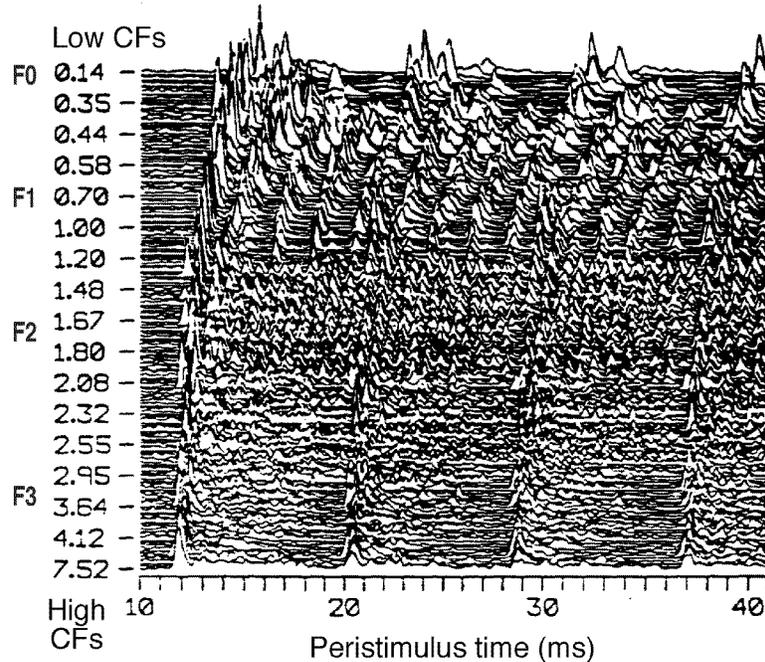


Figure 4.5. Responses of cat auditory nerve fibers to a multiformant vowel (/da/). Vertical axis: fiber CF (kHz); horizontal axis: peristimulus time (ms). Dominant harmonics in the respective formant regions impress their temporal structure on swaths of fibers, thereby partitioning cochlear (CF) territories. The production of intervals associated with the different partials thus reflects their relative magnitudes. From Seecker-Walker & Searle (1990), who analyzed data from Young & Sachs (1979).

magnitudes of the various partials. This is illustrated in the ANF responses depicted in Figure 4.5, where dominant harmonics (associated with vowel formants) drive different CF territories. As a consequence, when all of the intervals are summed together into the population-interval distribution, the distribution resembles the autocorrelation function of the stimulus (for frequencies up to the limits of phase-locking). Place per se is therefore not strictly necessary to encode stimulus spectrum -- each vowel with its characteristic formant structure and power spectrum produces a characteristic population-interval representation (that is precise and level-invariant). It is therefore possible and desirable to collapse correlogram-like representations across frequency in order to use the second dimension for time.

The result is an *autocorrelogram* (Fig. 4.6), which depicts the running autocorrelation of the stimulus. Although it is similar to a spectrogram in many respects, the autocorrelogram prominently depicts the dominant periodicities of speech sounds, i.e. voice pitch. We believe that effective front-end representations for speech analysis should represent those stimulus dimensions that are most prominent in speech perception, which we take to be voice pitch

(dominant periodicity), vowel quality (spectrum), fast amplitude and frequency patterns (dynamic aspects of timbre), and the manner in which they are grouped together to form discrete voices and objects. In contrast with spectrograms, the autocorrelogram effectively depicts invariants associated with voice pitch, which is crucial for speaker separation and enhancement.

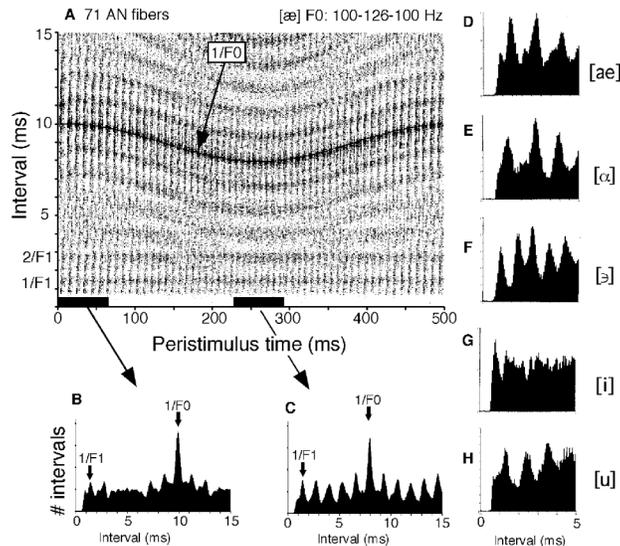


Figure 4.6. Population-interval representations of five synthetic vowels compiled from responses of 50-100 cat auditory nerve fibers. A. Population autocorrelogram for a variable-F₀ vowel /æ/, F₀= 100-126 Hz. The dark interval band closely follows the fundamental period. B-C. Cross sections of the autocorrelogram are population-interval distributions for 60 ms stimulus segments. The largest interval peak corresponds to voice pitch, while patterns of short intervals (< 5 ms) reflect dominant harmonics (formant structure). D-H. Population-interval distributions for five synthetic vowels, averaged over F₀s 100-126 Hz.

5 PHASE-EFFECTS IN F₀-BASED SEPARATION

It thus appears to us that the vast majority of the information that is utilized for perception of music and speech is *temporal in nature*. If we take this observation seriously, then new ways of thinking about the nature of auditory objects immediately present themselves. The ubiquity of phase-locking in early auditory processing means that spike timing patterns reflect the fine time structure (and phase structure) of the stimulus. This information is therefore potentially available for scene analysis mechanisms.

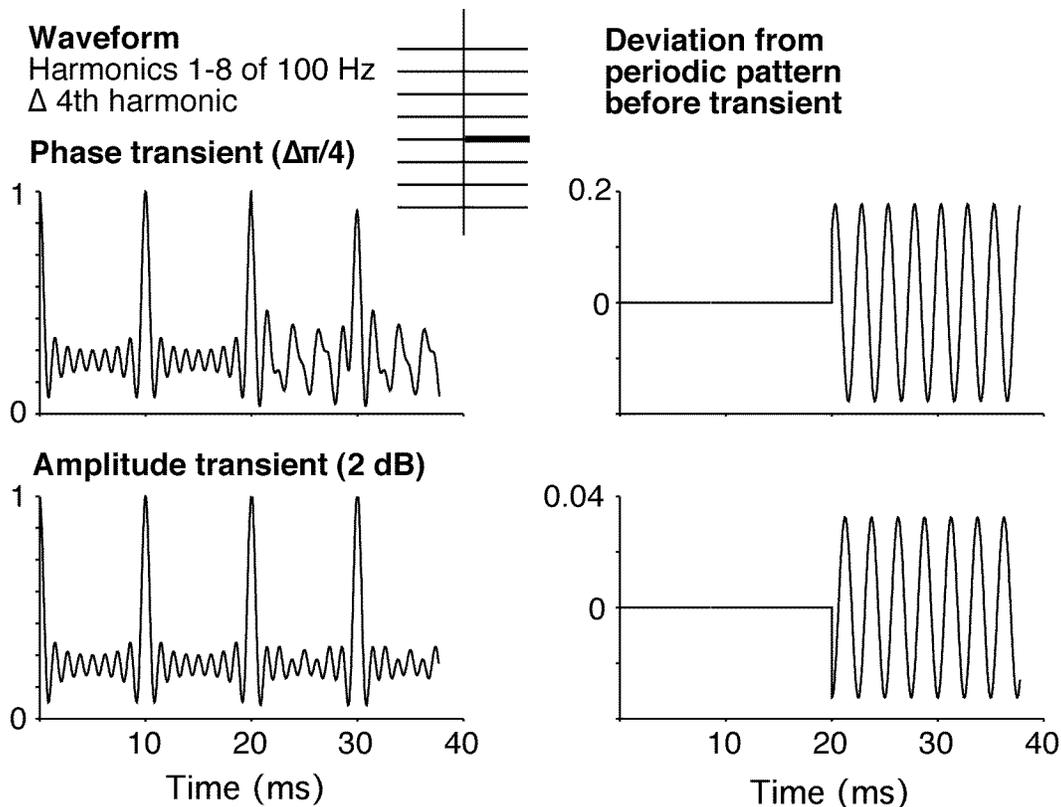


Figure 7. Phase and amplitude transients that cause individual harmonics to temporarily "pop-out" of harmonic complexes. Left. Waveforms of harmonic complexes having abrupt phase and amplitude shifts at 20 ms. Right. Subtraction of last two periods of the waveform from the first two (period-by-period deviation from perfectly periodic waveforms).

Although perception of pitch and timbre of stationary lower-frequency stimuli (the psychophysicists would say of resolved harmonics) is famously phase-insensitive, auditory grouping mechanisms are highly sensitive to abrupt *phase changes*. Harmonic complexes with F0s more than 2 semitones separate easily into two auditory objects with two distinct low (F0) pitches and timbres. We shall look at the double vowel case in more detail momentarily. Kubovy has demonstrated that abrupt changes in the phase and/or amplitude of a harmonic is sufficient to cause that harmonic to "pop-out". Each of these examples, including the separation of mistuned harmonics from complexes (see (Darwin and Gardner, 1986; Kubovy, 1981)), suggests a separation mechanism based on a period-by-period comparison. A mechanism that compared the waveform (spike pattern) of a preceding period with that of a subsequent period would register an ongoing disparity, and the form of this disparity would be that of the pure tone component. Assuming that pure tones are encoded temporally, then the auditory CNS would interpret this spike pattern disparity as a pure tone.

Double vowels with same and different fundamental periods would presumably be analyzed by this mechanism. The psychophysics of their separation and identification has been studied extensively (Assmann and Summerfield, 1990; Summerfield and Assmann, 1991). Models had been

proposed (de Cheveigne, 1999; Meddis and Hewitt, 1992), and their neural responses had been investigated (Cariani, 1995; Cariani and Delgutte, 1993; Palmer, 1988; Palmer, 1992). When the two vowels have the same F0, they fuse into one object; when the two F0s differ by a semitone or more, the vowels separate. When the F0s are the same, listeners are able to correctly identify both individual vowel constituents about half the time. For F0 separations of more than a semitone, listeners improve by 15-20%. A mechanism that built up recurrent temporal patterns would behave in a qualitatively similar way, fusing the two patterns when they have the same fundamental period and separating them when they have different ones.

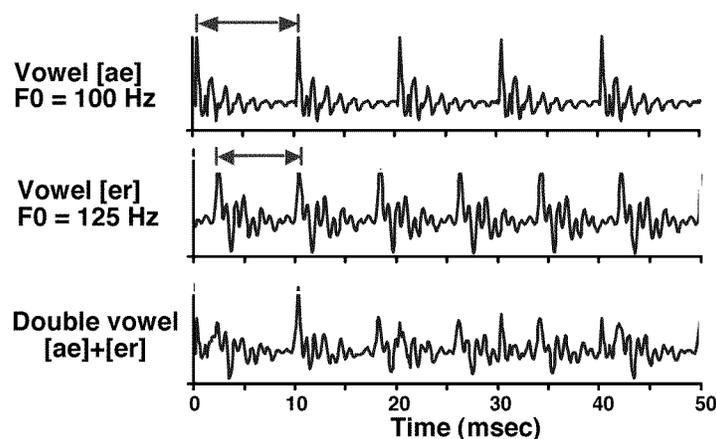


Figure 4.8. Double vowel stimuli.

A visual analogy can be constructed by overlaying two transparencies with arbitrary abstract patterns on them (e.g. random dots). When the transparencies are moved together, the patterns fuse, when they are moved independently, their respective patterns immediately separate. One intuitively feels that there is a mechanism that builds up the respective invariant patterns of dots in the transparencies and integrates these images when they are translated spatially. The mechanism will treat the two transparencies as one when the relations of dots on each sheet is stable relative to the dots on the other. Likewise, it will find two sets of invariant relations when the sheets are moved relative to each other -- while relations between dots on the two transparencies are rendered unstable and varying, those within a given sheet remain stable. In the auditory case, the relations are between patterns of spikes associated with the two vowels. The conventional visual scene analysis explanation assumes representation of the patterns as sets of active feature detectors (e.g. motion detectors, bug detectors), the scene analysis task being to separate the features associated with the two dot patterns on the transparencies. In the double vowel example two vowels separated by a few semitones in frequency drive the same neuronal frequency channels in the auditory nerve. The spike timing information that encodes the two vowels is temporally multiplexed in the same auditory nerve fibers (Figure 4.9). This suggests that the auditory system must use phase/fine timing information to separate the two vowels. Back to the visual case, since visual neurons also "phase-

lock" to moving images, it is not out the realm of possibility that the visual system utilizes spatial patterns of temporally-correlated spikes to represent and separate forms in an analogous manner (Cariani, 2004, in press). This temporal correlation hypothesis potentially explains why motion is essential for vision: why visual forms disappear when their positions are stabilized on the retina.

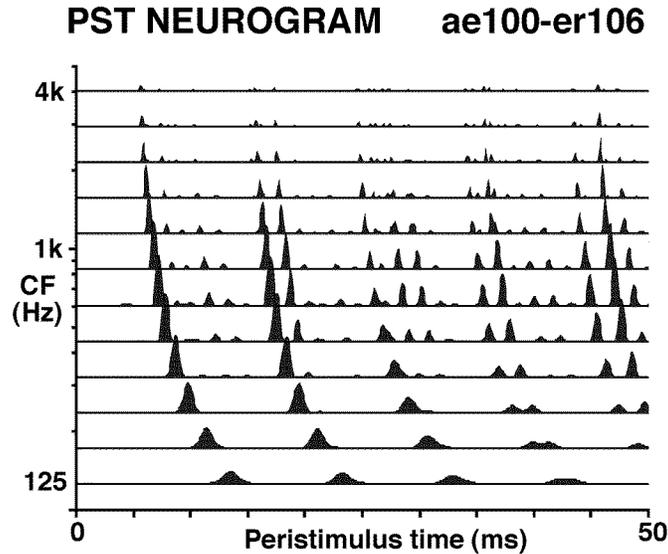


Figure 9. Simulated auditory nerve response to a double vowel with F0s separated by a semitone (6%). Information associated with the respective vowels is multiplexed in the time domain. The situation is similar for larger F0 separations.

One possibility then is that auditory objects are first formed via a low-level phase-sensitive mechanism that builds up periodic patterns and separates divergent ones. Such a mechanism would operate on the temporal coherence of fine time patterns (internal phase coherence) rather than detecting discrete features and sorting channels on that basis. Once objects were formed by such a mechanism, they would subsequently be analyzed by an phase-insensitive mechanism (pitch, timbre, loudness, location) . Recurrent neural timing nets are an attempt to demonstrate how this strategy for scene analysis could be implemented neurally.

We strongly believe that new information-processing models that focus on functional principles are absolutely essential for long-term progress in auditory neuroscience and speech recognition. In auditory neuroscience, it seems likely that computational biophysical simulations aimed at accounting for the input-output behavior of particular neuronal elements will not lead us to the functional principles we need to understand how the auditory system works as an information-processing system. In lieu of strong functional hypotheses, the neurophysiologists will be left to creating ever larger catalogs of neural responses. Audio separation and speech recognition should pay attention to natural auditory systems, if only because these are still by far the most effective sound processing and speech recognition devices on the planet. Again the attention needs to focus

on essential functional principles - which aspects of auditory function produce the performances that we see.

To date, there have been only a few isolated attempts to exploit the power of temporal codes in providing high-quality general-purpose auditory front-ends for speech recognizers (Ghitza, 1988; Ghitza, 1992). In the last decade correlograms have been used effectively to label and separate subsets of frequency channels (Wang and Brown, 1999). While these efforts, in incorporating fine timing information, represent a great improvement over purely spectral analyses, it seems to us unlikely that the auditory system operates by binding together discrete frequency channels. When we contemplate the image of spike activity in a simulated auditory nerve (Fig. 9) that uses broader, more physiologically realistic cochlear filters rather than much narrower "auditory filters" derived from (whole system) psychophysics, it seems immediately obvious that there are no discrete, narrowly-tuned neural frequency channels to be sorted. The situation is no better at higher levels of processing. Yet to be produced is neurophysiological evidence suggesting that two low-frequency pure tones 20% apart in frequency (easily heard as two discrete tones) will excite separable populations of neurons in higher auditory stations. The means by which the central auditory system achieves fine low frequency discrimination and selectivity thus remains a completely open (and largely neglected) question for auditory neuroscience. It is difficult to envision how investigations of scene analysis mechanisms at the cortical level can proceed without some prior understanding of the precise nature of neural representations of periodicity and spectrum at that level. On the other hand, well constructed studies of neural scene analysis mechanisms (e.g. Sinex, 2002) could conceivably shed precious light on the nature of these representations at higher centers.

In our view, temporal coding is not simply a special-purpose "hack" to do F0-based scene analysis, but a fundamental organizing principle of vertebrate auditory systems. Information processing in the auditory system may well be based on fine timing information that our present spectrographic representations throw away at an early stage. If neural analysis of sound primarily involves time-domain operations then it behooves us to explore similar strategies for sound analysis in artificial systems. This line of thinking suggests wider and more general use of correlograms and autocorrelograms for sound separation, front-end representations, and even back-end recognition strategies. Ideally we should use such representations in conjunction with processing strategies that parallel human auditory scene analysis. Here the most important mechanisms group sounds by temporal contiguity (common onset/offset) and by common harmonicity (F0). When we have both auditory representations and sound separation strategies that exploit the temporal microstructure of sounds, then we should be well on our way to developing much more robust bottom-up automatic speech recognition systems.

6 RECURRENT TIMING NETS

For these and other reasons, we have strived to develop new heuristics for how auditory images might be formed and separated. Both feedforward and recurrent networks we have been considered and their basic computational

properties were explored (Cariani, 2001a; Cariani, 2001b; Cariani, 2004 (in press)). Neural timing nets demonstrate how analog time-domain filtering operations could conceivably be performed in neural network implementations. They expand the realm of possible signal-processing mechanisms available to nervous systems. We hope they will have the effect of catalyzing new functional hypotheses for how information could be represented, transmitted, multiplexed, broadcast, analyzed, and integrated. These networks have also been investigated in the context of music perception (Cariani, 2002), as possible approaches to tonality and rhythmic induction.

Recurrent timing networks were inspired in different ways by models of stabilized auditory images (Patterson et al., 1995), neural loop models (Thatcher and John, 1977), adaptive timing nets (MacKay, 1962), adaptive resonance circuits (Grossberg, 1988), the precision of echoic memory, and the psychology of temporal expectation (Jones, 1976; Miller and Barnet, 1993). Although much is known about time courses of temporal integration that are related to auditory percepts (pitch, timbre, loudness, location, object separation, and various masking effects), we currently have few good models for how incoming information in the auditory periphery is integrated over time by the central auditory system to form stabilized auditory percepts. If the information involved is indeed temporally-coded, then architectures that store temporal patterns in reverberating circuits eventually come to mind. One envisions the signals themselves circulating in closed transmission loops or regenerated via cellular recovery mechanisms. These temporal memory traces (temporal echoic memories) would be compared with incoming patterns via coincidence-detectors that compute temporal correlations. Neural representations would thus build up over time, dynamically creating sets of perceptual expectations that could either be confirmed or violated. Periodic signals, such as rhythms, would thereby create strong temporal expectancies (Cariani, 2002; Fraisse, 1978).

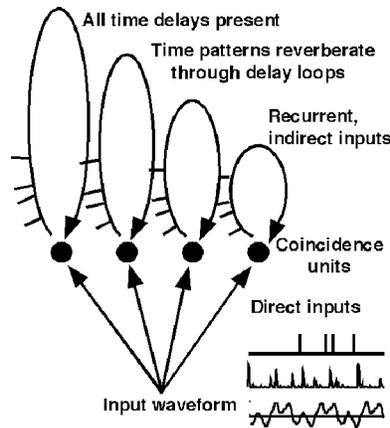


Figure 10. Simple recurrent timing net.

The simplest recurrent timing networks imaginable consist of a 1-D array of coincidence detectors having common direct inputs (Figure 10). The output of each coincidence element is fed into a recurrent delay line such that the output of

the element at time t circulates through the line and arrives τ milliseconds later (the signal that arrives back at time t is the one that was emitted at $t - \tau$). A processing rule governs the interaction of direct and circulating inputs.

In their development the networks have evolved from simple to more complex. In the first simulations (Cariani, 2001a), binary pulse trains (resembling spike trains) with repeated, randomly selected pulse patterns (e.g. 100101011-100101011-100101011...) were passed through the network. For each time step, incoming binary pulses were multiplied by variable-amplitude pulses arriving through the delay loop. In the absence of a coincidence with a circulating pulse, the input pulse was fed into the delay loop without facilitation. When coincidences between incoming and circulating pulses occurred, the amplitude of the circulating pulse was increased by 5% and the pulse was fed back into the loop. It was quickly realized that such networks rapidly build up any periodic pulse patterns in their inputs, even if these patterns are embedded amidst many other pulses. A periodic pattern invariably builds up in the delay loop whose recurrence time matches its repetition time. Thus, recurrent time patterns are repeatedly correlated with themselves to build up to detection thresholds. In effect, these autocorrelating loops dynamically create matched filters from repeating temporal patterns in the stimulus. In this manner, temporal-pattern invariances are enhanced relative to uncorrelated patterns – the network functions as a pattern-amplifier. When two repeating temporal patterns each with its own repetition period were summed together and presented to such nets, the two patterns emerged in the two different delay loops that had recurrence times that corresponded to the repetition periods of the patterns. Although the proportional facilitation rule distorted signal amplitudes, the temporal patterns of pulses corresponding to the two rhythms could be recovered in the circulating waveforms. A neural network can therefore carry out an analog-style separation of signals in the time-domain. To do this, inputs need to be temporally coded, processing elements must have sufficiently narrow coincidence windows, delays must be relatively precise, and processing rules must be judiciously chosen.

7 SEPARATION OF DOUBLE VOWELS

Although binary pulse trains resemble spike trains of individual neurons, most real neural information processing appears to be carried out by large ensembles of neurons working in concert. Subsequent simulations (Cariani, 2001a) therefore used positive real-valued input signals that qualitatively resemble neural post-stimulus time histograms (e.g. time series of spike counts that would be produced by an ensemble of similar neural elements whose discharges were stimulus-locked). Proportional facilitation was replaced by a processing rule that adaptively adjusted the output signal in a more graceful and less distorting manner. Multichannel implementations subsequently processed the double vowels using an auditory nerve front-end with 24 CF channels. The instantaneous spike rate of each frequency channel was fed into an array of delay loops and the autocorrelations of the circulating waveforms in corresponding delay loops were combined to produce simulated population-interval distributions.

Our current single-channel implementation uses a simple error-adjustment processing rule that can operate on signals with both positive and negative values. For each loop with recurrence time τ , at time t , $X(t)$ is the direct input signal, $H(t-\tau)$ is the incoming circulating signal, and $H(t)$ is the outgoing circulating signal. $H(t) = H(t-\tau) + B_{\tau}[X(t)-H(t-\tau)]$ describes the input-output function of each processing element. B_{τ} determines the rate of adjustment, and its dependence on τ ($B_{\tau} = \tau/33 \text{ ms}$) ensures that shorter loops are not favored.

To a signal processing engineer, the net somewhat resembles a temporally-coded neural implementation of a bank of comb filters, albeit ones with very short (1-2 period) temporal integration times. It would be a mistake to dismiss them as simple autocorrelations (in the same way that it would be a mistake to reduce all the refinements of spectral analysis to Fourier's Theorem). All of the most effective strategies for implementing F0-based separations (comb filters, correlograms, cancellation operations) are in one way or another formally related to autocorrelation. But with their short memories, time-domain implementations, and avoidance of early windowing, these nets have more in common with the sample-by-sample and period-by-period harmonic cancellation strategies of de Cheveigne (Chapter 16, this volume) than they do with traditional sharp comb filters that utilize long integration times.

Synthetic, three-formant double vowels (/ae/, /er/) with different fundamentals (100, 112 Hz) were summed and processed by the network (Fig. 4.11). The signals circulating in the 150 delay loops are shown in the response map, where it can be seen that the recurrence times of the loops with the highest average signal strength correspond to the periods of the two vowels (8.9 and 10 ms). The signals circulating in these two delay channels after 70 ms of processing highly resemble the two vowel constituents. Correlations between the autocorrelations of these processed signals and those of the individual vowels show how the signal separation unfolds over processing time.

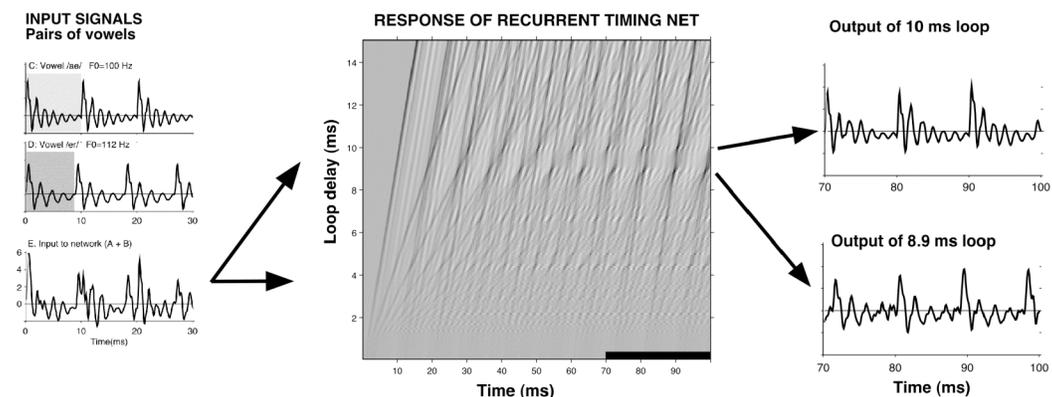


Figure 4.11. Separation of a double vowel (ae-er) with different F0s (100, 112 Hz) into its constituent waveforms by means of a recurrent timing net. Top. Array of processing elements and delay loops. Individual synthetic vowels and their summed waveforms. Bottom. Response of the timing network to the double

vowel. Arrows emerge from the two loops with the biggest signals. Waveforms circulating in the two loop channels after 70 ms.

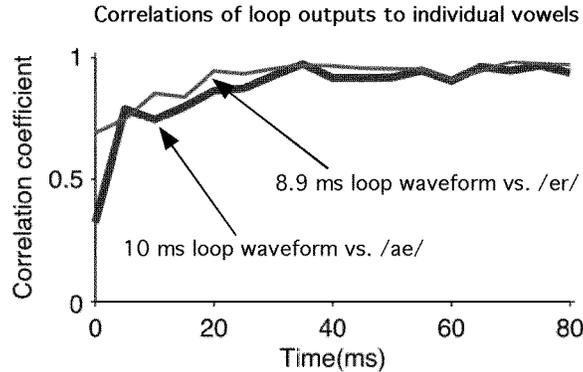


Figure 4.12. Performance of the network in separating the vowels. Degree of similarity between the output waveforms and the individual vowel constituents, as measured by the correlation between their autocorrelations. The network rapidly separates the two vowels, within 2-3 periods.

In both single- and multi-channel cases, when vowel fundamentals were separated by a semitone or more, the autocorrelations (and hence, power spectra) of the constituent vowels could be accurately recovered. The quality of the separations improved as a function of ΔF_0 and vowel duration. The multichannel simulation demonstrated how recurrent timing nets could be scaled up to process multichannel positive, real-valued signals not qualitatively unlike those produced by auditory nerve arrays. It also showed that auditory objects can be separated when they activate the same sets of broadly tuned frequency channels (i.e. without usable rate-place information), provided that phase-locked fine timing information is available. The information related to multiple auditory objects (the two vowels) is embedded in the phase structure of the stimulus and phase-locked neural responses. This is a relatively straightforward auditory example of how information can be multiplexed in the time domain. The networks also demonstrate how an auditory scene analysis system could exploit phase-coherence and F_0 -differences without first carrying out explicit estimations of F_0 and segregating frequency channels on that basis. For example, in Wang & Brown (1999), correlograms (f, τ) label frequency channels with common F_0 -related autocorrelation profiles, which are then grouped using an array of synchronizing oscillators. In timing networks, formation of auditory objects occurs prior to analyses that yield object attributes (F_0 pitch, timbre).

One can ask how well these networks handle more than two auditory objects. A third vowel /ee/ with yet a different fundamental (125 Hz) was added to the mixture. This is akin to the problem of hearing out three different kinds of musical instruments playing different notes (first that there are three different notes, second that instruments with different timbres are playing the three notes). Processing by the network resulted in the appearance of another strong signal in the response map. Separation of the signals in the three-vowel case was somewhat

slower than for two vowels, but there was only a slight reduction in the final quality of the separated signals. The performance of the network therefore appears to be highly robust.

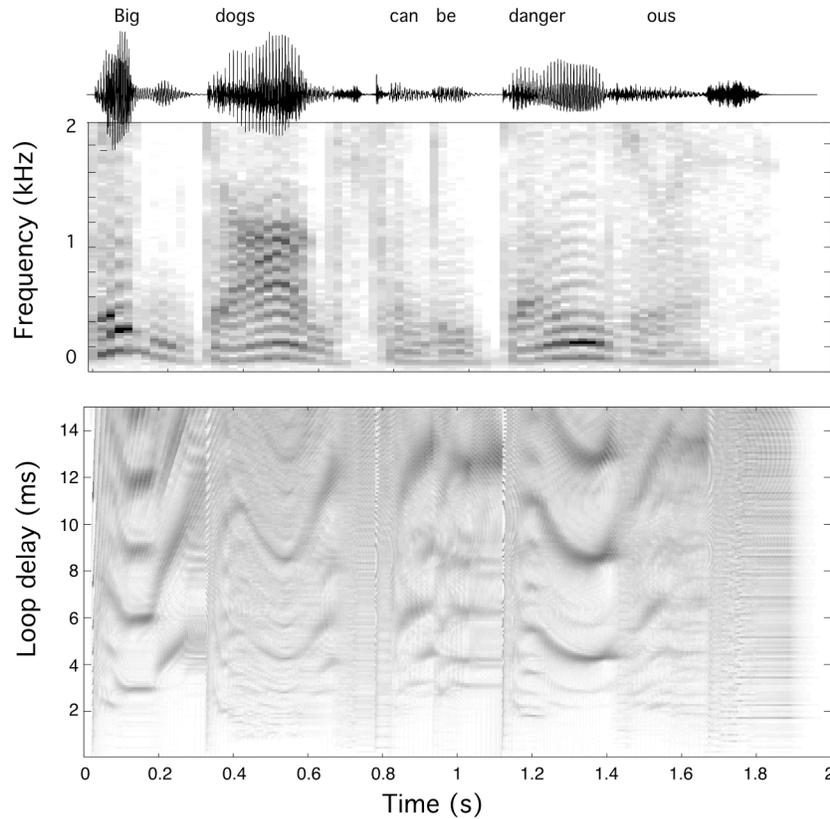


Figure 13. Response of a recurrent timing net to running speech. Top plots. Waveform and spectrogram for HINT sentence, "Big dogs can be dangerous". Bottom. Output of RTN after smoothing and normalization. Contours related to the voice pitch of the male speaker are seen in loop channels near 10 ms.

We are developing RTN-based voice pitch trackers that can handle running speech (Fig. 4.13). The general form of the RTN response looks somewhat similar to the stimulus autocorrelogram, and the RTN has no difficulty following the rapidly varying pitch contour of the speaker. F0 tracks can be constructed over multiple time scales (samples, loop periods, syllables, sentences) and the signals in the corresponding loop-time trajectories can be easily and relatively seamlessly assembled on a sample-by-sample basis without resorting to windowing and without leaving the time domain. Although F0-tracking only separates voiced segments, voicing provides a temporal framework for delimiting unvoiced segments. If voiced segments of different speakers have asynchronous onsets and offsets, it may be possible to use continuity rules to assign unvoiced segments to different speakers. Once voiced segments can be identified, then adjacent voiced-

unvoiced or unvoiced-voiced patterns may also be analyzable as well-formed CV units.

8 ENHANCEMENT OF VOWELS IN NOISE

A possible use of recurrent timing nets is for enhancement of periodic sounds in noisy environments. Such processing would be useful for processing music and voiced speech. Reductions in effective S/N ratios could be expected to improve speech reception by human listeners and automatic recognition by machines. Related kinds of correlation-based strategies were used in the 1950's to detect periodic signals in noise (Lange, 1967; Meyer-Eppler, 1953), in situations where the period of the target signal was known *a priori*. The present networks systematically sample all possible delays, such that the optimum delay(s) can be determined by choosing the loop(s) with the largest signal rms.

In order to assess the performance of the network in noise, a synthetic, three-formant vowel (/ae/, F0=100 Hz, S/N = -20-20 dB) was added to frozen white noise at different S/N ratios that ranged from -20 to 20 dB. The input and output signals from the optimum delay loop (tau = 10 ms) are shown in the top panels of Figure 6. The bottom panels show the correlation between the autocorrelation of these signals and that of the vowel in near-quiet (20 dB S/N). Similarities between the processed signals and the minimal-noise case improve with S/N and processing time. For all S/N ratios less than 1, the network produced output signals (thick curves) that had higher correlations than the unprocessed, input signals (thin curves). Processing by the network shifts the curves to the left, an improvement in S/N by roughly 4-10 dB that is comparable to improvements that have been reported using comb filters (Stern, 2003).

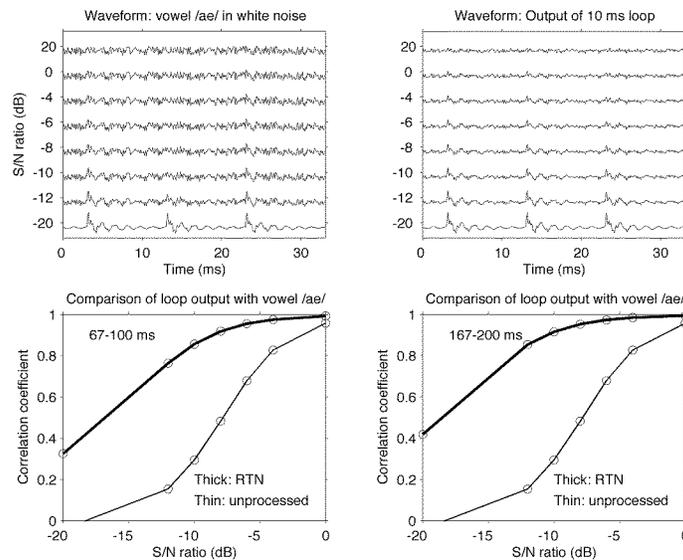


Figure 14. RTN representation of a vowel in white noise.

9 THEORETICAL CONSIDERATIONS

We have been contemplating some of the longer-range theoretical implications that relate to brains as general-purpose self-organizing correlation machines that extract invariant patterns in their inputs. Recurrent timing nets use the periodic patterns in their inputs to dynamically form matched templates that they compare with subsequent inputs. For ease of visualizing their behavior, we have considered ordered arrays of monosynaptic delay loops. It is conceivable, however, that such processing can also be carried out in randomly connected networks, provided that recurrent, multisynaptic pathways are available that span a wide range of loop-delays. If such networks use coincidence elements that are transiently facilitated by temporal coincidences, then it is not hard to envision how they might support dynamically-formed reverberatory memories capable of retaining temporal patterns and interspike interval statistics. In their operation such networks would be akin to self-organizing recurrent synfire chains (Abeles, 1990, 2004) in which both synchrony and temporal patterning of spikes play critical roles. Processing using spike statistics "liberates signals from the wires", since the respective identities of signals then no longer depends on which particular input lines the respective signals conveyed. It opens the possibility of more flexible kinds of neural networks that can multiplex signals, broadcast them, and recombine them in novel ways.

ACKNOWLEDGMENTS

We thank Pierre Divenyi for his encouragement. This work was supported in part by the National Science Foundation, NSF-EIA-BITS-013807

Abbreviations used

CF: characteristic frequency, a feature of the frequency tuning properties of auditory neurons. CF is the pure tone frequency to which an auditory neuron is most sensitive. More precisely, CF is the frequency of the pure tone with the lowest sound pressure level that reliably causes the neuron to fire at a rate higher than its resting, "spontaneous" rate (in the absence of a tone). Operationally, this is often defined as a change in firing rate > 20 spikes/sec above the average spontaneous rate. The CF of an auditory nerve fiber reflects the cochlear location and consequent frequency tuning of the inner hair cell from which the fiber receives its synaptic inputs.

CNS: central nervous system

SPL: sound pressure level, the ratio of average sound pressure (root-mean-square (rms) of instantaneous pressure $P(t)$) to the reference pressure level P_{ref} of 20 micro Pascals (.0002 dynes/cm²). dB SPL: sound pressure level, expressed in decibels, i.e. $SPL = 20 * \log(rms(P(t))/P_{ref})$

2AFC ("Two-alternative forced choice"): Psychoacoustics experimental paradigm in which a subject is forced to choose between two alternative perceptual judgments, e.g. whether one sound is higher or lower in pitch, or whether two sounds are the same or different.

LGN: lateral geniculate nucleus, the thalamic nucleus of the ascending visual pathway.

F0: fundamental frequency, the greatest common integer divisor of a harmonic series, or alternately, the repetition frequency of a complex temporal pattern. For example, 100 Hz is the

fundamental frequency of the summation of multiple, harmonically-related pure tones having frequencies of 300, 400, 500, and 600 Hz. 100 Hz would also be the fundamental frequency of any periodic waveform which repeats every 10 ms (but no sooner). The voice pitches of human speakers are heard at the fundamental frequencies of voiced speech segments. Within frequency and intensity limits, almost all periodic sounds evoke a pitch at their fundamental.

REFERENCES

- Abeles, M. (2004). Time is precious. *Science* **304** (23 April): 523-524.
- Abeles, M., 1990. *Corticonics*. Cambridge University Press, Cambridge.
- Assmann, P.F. and Summerfield, Q., 1990. Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies. *J. Acoust. Soc. Am.*, 88: 680 - 697.
- Boring, E.G., 1942. *Sensation and Perception in the History of Experimental Psychology*. Appleton-Century-Crofts, New York.
- Bregman, A.S., 1981. Asking the "what for" question in auditory perception. in, *Perceptual Organization* M. Kubovy and J.R. Pomerantz (Editors). Lawrence Erlbaum Assoc., Hillsdale, NJ, pp. 99-118.
- Bregman, A.S., 1990. *Auditory Scene Analysis: The Perceptual Organization of Sound*. MIT Press, Cambridge, MA, 773 pp.
- Burns, E.M., 1999. Intervals, scales, and tuning. in, *The Psychology of Music (2nd Ed.)* D. Deutsch (Editor) Academic Press, San Diego, pp. 215-264.
- Cariani, P., 1995. As if time really mattered: temporal strategies for neural coding of sensory information. *Communication and Cognition - Artificial Intelligence (CC-AI)*, 12(1-2): 161-229 (Reprinted in: K Pribram, ed. *Origins: Brain and Self-Organization*, Hillsdale, NJ: Lawrence Erlbaum, 1994; 208-252.).
- Cariani, P., 1999. Temporal coding of periodicity pitch in the auditory system: an overview. *Neural Plast*, **6**(4): 147-72.
- Cariani, P., 2001a. Neural timing nets. *Neural Networks*, 14(6-7): 737-753.
- Cariani, P., 2001b. Neural timing nets for auditory computation. in, *Computational Models of Auditory Function* S. Greenberg and M. Slaney (Editors). IOS Press, Amsterdam, pp. 235-249.
- Cariani, P., 2002. Temporal codes, timing nets, and music perception. *J. New Music Res.*, 30(2): 107-136.
- Cariani, P., 2004 (in press). Temporal codes and computations for sensory representation and scene analysis. *IEEE Transactions on Neural Networks, Special Issue on Temporal Coding for Neural Information Processing*.
- Cariani, P. and Delgutte, B., 1993. Interspike interval distributions of auditory nerve fibers in response to concurrent vowels with same and different fundamental frequencies. *Assoc. Res. Otolaryngology. Abs.*: 373.
- Darwin, C.J. and Gardner, R.B., 1986. Mistuning a harmonic of a vowel: grouping and phase effects on vowel quality. *J Acoust Soc Am*, 79(3): 838-45.
- de Cheveigne, A., 1999. Waveform interactions and the segregation of concurrent vowels. *J Acoust Soc Am*, 106(5): 2959-72.
- de Cheveigne, A., 2004. The cancellation principle in acoustic scene analysis. In: *Perspectives on Speech Separation*, P. Divenyi (Editor), Kluwer, New York, pp. 241-255.
- de Cheveigné, A., 2004, in press. Pitch perception models. in, *Pitch*, C.J. Plack and A.J. Oxenham (Editors). Springer Verlag, New York.
- Fraisse, P., 1978. Time and rhythm perception. in, *Handbook of Perception. Volume VIII. Perceptual Coding* E.C. Carterette and M.P. Friedman (Editors). Academic Press, New York, pp. 203-254.
- 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.

- Ghitza, O., 1992. Auditory nerve representation as a basis for speech processing. in, *Advances in Speech Signal Processing* S. Furui and M.M. Sondhi (Editors). Marcel Dekker, New York, pp. 453-485.
- Grossberg, S., 1988. *The Adaptive Brain, Vols I. and II.* Elsevier, New York.
- Handel, S., 1989. *Listening: An Introduction to the Perception of Auditory Events.* MIT Press, Cambridge, 597 pp.
- Hartmann, W.M., 1988. Pitch perception and the segregation and Integration of auditory entities. in, *Auditory Function: Neurobiological Bases of Hearing* G.M. Edelman (Editor) John Wiley & Sons, New York, pp. 623-347.
- Irvine, D.R.F., 1986. *The Auditory Brainstem.* Progress in Sensory Physiology 7. Springer-Verlag, Berlin, 279 pp.
- Jones, M.R., 1976. Time, our lost dimension: toward a new theory of perception, attention, and memory. *Psychological Review*, 83(5): 323-255.
- Kim, D.O. and Molnar, C.E., 1979. A population study of cochlear nerve fibers: comparison of spatial distributions of average-rate and phase-locking measures of responses to single tones. *J. Neurophysiol.*, 42(1): 16-30.
- Kubovy, M., 1981. Concurrent-pitch segregation and the theory of indispensable attributes. in, *Perceptual Organization* M. Kubovy and J.R. Pomerantz (Editors). Lawrence Erlbaum Assoc., Hillsdale, NJ, pp. 55-98.
- Lange, F.H., 1967. *Correlation Techniques.* Van Nostrand, Princeton, 464 pp.
- MacKay, D.M., 1962. Self-organization in the time domain. in, *Self-Organizing Systems 1962* M.C. Yovitts, G.T. Jacobi and G.D. Goldstein (Editors). Spartan Books, Washington, D.C., pp. 37-48.
- Meddis, R. and Hewitt, M.J., 1992. Modeling the perception of concurrent vowels with different fundamental frequencies. *J. Acoust. Soc. Am.*, 91: 233-245.
- Mellinger, D.K. and Mont-Reynaud, B.M., 1996. Scene analysis. in, *Auditory Computation* H. Hawkins, T. McMullin, A.N. Popper and R.R. Fay (Editors). Springer Verlag, New York, pp. 271-331.
- Meyer-Eppler, W., 1953. Exhaustion methods of selecting signals from noisy backgrounds. in, *Communication Theory* W. Jackson (Editor) Butterworths, London, pp. 183-194.
- Miller, R.R. and Barnet, R.C., 1993. The role of time in elementary associations. *Current Directions in Psychological Science*, 2(4): 106-111.
- Palmer, A.R., 1988. The representation of concurrent vowels in the temporal discharge patterns of auditory nerve fibers. in, *Basic Issues in Hearing* H. Duifhuis, J.W. Horst and H.P. Wit (Editors). Academic Press, London, pp. 244-251.
- Palmer, A.R., 1992. Segregation of the responses to paired vowels in the auditory nerve of the guinea pig using autocorrelation. in, *The Auditory Processing of Speech* S. M.E.H. (Editor) Mouton de Gruyter, Berlin, pp. 115 - 124.
- Patterson, R.D., Allerhand, M.H. and 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.
- Rose, J.E., Hind, J.E., Brugge, J.R. and 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.
- Secker-Walker, H.E. and Searle, C.L., 1990. Time-domain analysis of auditory-nerve-fiber firing rates. *J. Acoust. Soc. Am.*, 88(3): 1427-1436.
- Sinex, D.G., H., Sabes, S.J. and Li, H., 2002. Responses of inferior colliculus neurons to harmonic and mistuned complex tones. *Hear. Res.*, **168**: 150-62.
- Stern, R. 2003. Signal separation motivated by auditory perception. *Perspectives on Speech Separation*, Montreal, October 30- November 1, 2003. <http://www.ebire.org/speechseparation>.
- Summerfield, Q. and 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. and John, E.R., 1977. *Functional Neuroscience, Vol. I. Foundations of Cognitive Processes*. Lawrence Erlbaum, Hillsdale, NJ, 382 pp.
- Wang, D.L. and Brown, G.J., 1999. Separation of speech from interfering sounds based on oscillatory correlation. *IEEE Trans. Neural Networks*, 10(3): 684-697.
- Young, E.D. and 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.

Outline of a cybernetic theory of brain function based on neural timing nets

Cybernetic
theory of brain
function

1219

Peter Cariani

*Hearing Research Center and Department of Biomedical Engineering,
Boston University, Boston, Massachusetts, USA*

Abstract

Purpose – The purpose of this paper is to outline an integrative, high-level, neurocomputational theory of brain function based on temporal codes, neural timing nets, and active regeneration of temporal patterns of spikes within recurrent neural circuits that provides a time-domain alternative to connectionist approaches.

Design/methodology/approach – This conceptual-theoretical paper draws from cybernetics, theoretical biology, neurophysiology, integrative and computational neuroscience, psychology, and consciousness studies.

Findings – The high-level functional organization of the brain involves adaptive cybernetic, goal-seeking, switching, and steering mechanisms embedded in percept-action-environment loops. The cerebral cortex is conceived as a network of reciprocally connected, re-entrant loops within which circulate neuronal signals that build up, decay, and/or actively regenerate. The basic signals themselves are temporal patterns of spikes (temporal codes), held in the spike correlation mass-statistics of both local and global neuronal ensembles. Complex temporal codes afford multidimensional vectorial representations, multiplexing of multiple signals in spike trains, broadcast strategies of neural coordination, and mutually reinforcing, autopoiesis-like dynamics. Our working hypothesis is that complex temporal codes form multidimensional vectorial representations that interact with each other such that a few basic processes and operations may account for the vast majority of both low- and high-level neural informational functions. These operational primitives include mutual amplification/inhibition of temporal pattern vectors, extraction of common signal dimensions, formation of neural assemblies that generate new temporal pattern primitive “tags” from meaningful, recurring combinations of features (perceptual symbols), active regeneration of temporal patterns, content-addressable temporal pattern memory, and long-term storage and retrieval of temporal patterns via a common synaptic and/or molecular mechanism. The result is a relatively simplified, signal-centric view of the brain that utilizes universal coding schemes and pattern-resonance processing operations. In neurophenomenal terms, waking consciousness requires regeneration and build up of temporal pattern signals in global loops, whose form determines the contents of conscious experience at any moment.

Practical implications – Understanding how brains work as informational engines has manifold long-reaching practical implications for design of autonomous, adaptive robotic systems. By proposing how new concepts might arise in brains, the theory bears potential implications for constructivist theories of mind, i.e. how observer-actors interacting with one another can self-organize and complexify.

Originality/value – The theory is highly original and heterodox in its neural coding and neurocomputational assumptions. By providing a possible alternative to standard connectionist theory of brain function, it expands the scope of thinking about how brains might work as informational systems.

Keywords Cognition, Consciousness, Neural nets, Cybernetics, Autopoiesis, Brain

Paper type Conceptual paper

Introduction

How brains operate as informational engines is still largely an unsolved mystery, and it is far from clear that current connectionist neural network theories can provide adequate explanations. This paper outlines an alternative, cybernetic theory of brain function based on temporal codes and computations.



Embedded goal systems within percept-action cycles

It has been long recognized by biologists and psychologists that animals perceive their current surrounds and act accordingly, usually appropriately, in successive on-going cycles of sensation, decision making, and action. Processes of deliberation, how an animal decides what to do in a given perceived situation, involve embedded goal systems, and anticipatory action-selection mechanisms that coordinate perceptions and actions.

Nervous systems that implement these coordinations can thus be regarded as cybernetic percept-action systems (Cariani, 2011), i.e. purposive systems whose behaviour is contingent on internal goals, anticipatory prediction and feedback control mechanisms, memory of prior experience, and current sensory inputs (Arbib, 1972; Sommerhoff, 1974; Powers, 1973). The highest level of feedback control switches between largely discrete behavioural modes such as fighting/fleeing/exploring/sleeping/eating/mating (Kilmer and McCulloch, 1969). Affect plays a pivotal role in switching between most of these modes (Barrett and Bliss-Moreau, 2009). Within each mode, various goal circuits compete for control of behaviour on the basis of goal priority (immediate urgency) and current degree of satisfaction (see Verschure *et al.*, 2014 for a contemporary model).

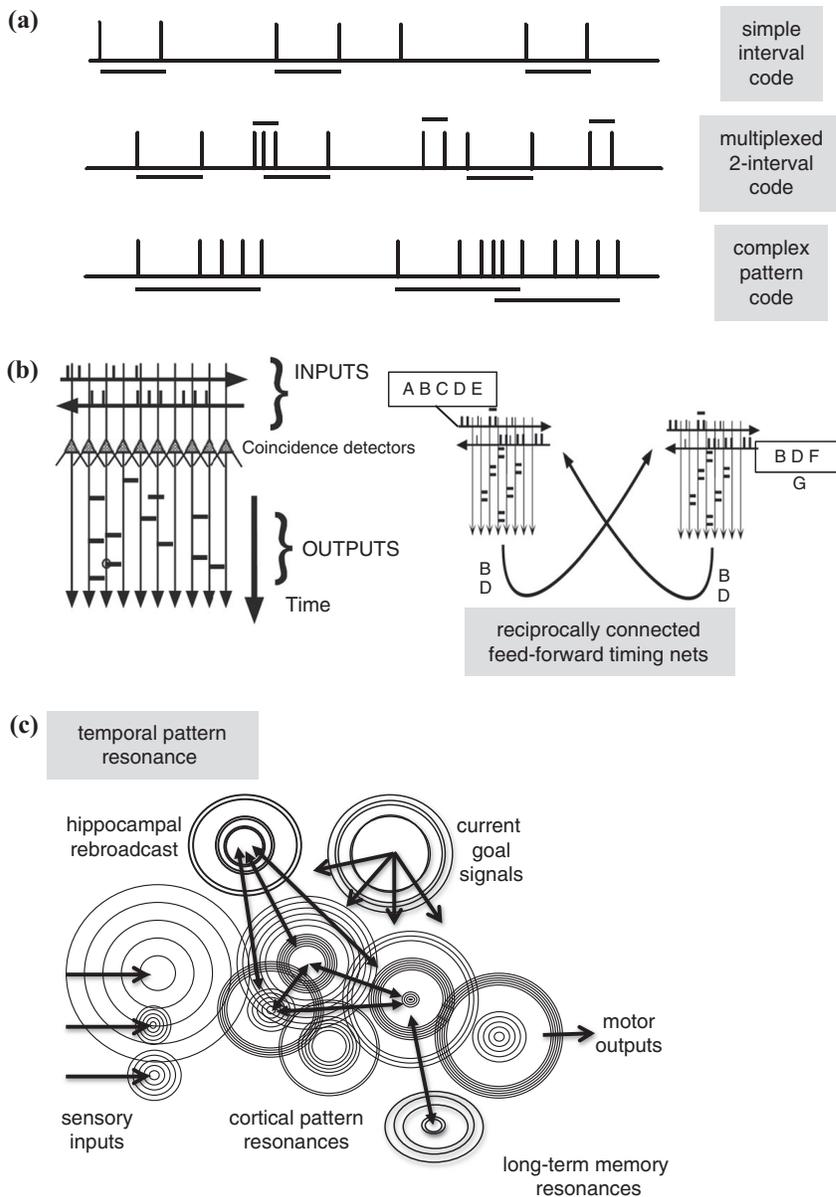
The basic structural plan of animal nervous systems (e.g. Swanson, 2012) consists of a relatively small number of component subsystems that subservise different types of informational functions: sensing (receptors), effecting (muscles), analysing sensory patterns (in mammals, primary and secondary sensory cerebral cortex), object-recognition (multimodal cortex), action-planning (premotor cortex), sequence generation (motor cortex), fine timing control (cerebellum), task-dependent steering systems for gating attention and motor switching (basal ganglia), spatial-relational processing (parietal lobe), regeneration of working memory and memory consolidation (hippocampus), and anticipatory reward prediction and correction systems (midline dopamine circuits). Autonomic systems regulate body functions, with interoceptive sensory organs providing information related to body states (satiety, pain). Limbic analyzers of global internal state provide general dispositions for action (emotions).

Temporal codes and computations

Informational distinctions are mediated through specific patterns of neural activity (neural codes Boring, 1942; Uttal, 1972, 1973; Rieke *et al.*, 1997; Cariani, 1995). In terms of internal, on-going dispositional states and overt behaviours, these are “differences that make a difference”. Neural codes can be divided into channel codes and temporal codes: in channel codes, activation of particular subsets of neurons (channels) convey distinctions (e.g. through across neuron firing rate profiles), whereas in temporal codes particular temporal patterns of spikes convey distinctions.

Connectionist theory adopts the assumption that the central codes operant in brains are firing rate channel codes, whereas the neural timing net theory outlined here posits that these are complex temporal pattern codes. Although channel coding has generally been adopted in the neurosciences as the conventional, default neural coding assumption, throughout the history of neuroscience temporal codes have always been proposed alongside them as alternatives (Boring, 1942). We have previously discussed many of the relative merits of these two kinds of coding schemes (Cariani, 1995, 1999, 2001, 2004).

In simple temporal codes, one temporal parameter conveys one perceptual distinction (Figure 1(a)). For example, in the early auditory system distributions of times between spikes (interspike intervals) produced by the same neurons subservise pitch perception (Cariani, 1995, 1999; Cariani and Delgutte, 1996). The pitch that is



Notes: (a) Simple and complex temporal codes; (b) left, feed-forward timing net (FFT); right, reciprocally connected FFTNs; (c) visual depiction of pattern-resonance processes, showing signal interactions, mutual reinforcement and competition, goal-driven facilitation, hippocampal rebroadcast, and interactions with long-term memory

Figure 1.
Temporal codes and
computations

heard corresponds to the most common set of related interspike intervals that are produced in the auditory nerve. Binaural auditory localization in the horizontal plane utilizes sub-millisecond spike timing differences produced by corresponding neurons in neural pathways that originate in the two ears. Analogous examples exist in nearly every sense modality (Cariani, 2001b).

Complex temporal pattern codes can be formed from combinations of simple temporal pattern primitives (Figure 1(a)). The appearance of a particular temporal pattern in an ensemble of neurons functions as an independent primitive feature. Combinations of temporal patterns that are present at any given time form a feature-vector. The various patterns can be interleaved or embedded in other patterns of spikes, permitting signal multiplexing (concurrent transmission of multiple types of signals over the same neuronal transmission lines). A complex temporal coding scheme that was proposed in the past for multiple cutaneous sensory qualities (Emmers, 1981) provides a concrete example of how such codes might be organized.

A neural code must convey two types of information – the type of distinction the signal conveys (e.g. pitch vs perceived location) and the attribute distinction itself (e.g. particular pitches or locations). In channel codes, the identity of the channel (which neuron, as determined by its place in the network, its interconnectivities) conveys informational type, whereas patterns of channel activations convey different attribute values. For channel-coding schemes, neural channel-identities maintained via specific interconnections are absolutely critical for function. A neural firing rate is meaningless to the rest of the system without the identity of the neuron that is firing. In contrast, because temporal schemes encode the type of the signal in the form of the message, highly specific interneural connectivities and signal transmission paths are no longer essential for function. Temporal codes thus permit “signals to be liberated from wires” and this in turn enables broadcast modes of signal distribution. Neuronal assemblies downstream can be selectively tuned such that they respond to particular temporal patterns embedded in their inputs. Broadcast, multiplexing, and selective tuning enable decentralized communications in which neural assemblies can respond only to those incoming signal patterns that are relevant to their functional roles. Some codes may be restricted to local neuronal populations (e.g. restricted cortical regions) that handle specific types of information, whereas others may be propagated more globally, re-broadcast by the hippocampus, and then consolidated into long-term memory.

Temporal codes with sub-millisecond precisions can be found in a wide range of sensory systems (Cariani, 2001b). In general, as one proceeds up ascending sensory pathways, away from receptor surfaces, stimulus-related fine timing information becomes successively less apparent, though not completely absent (Cariani, 1999). In the auditory system, despite some progress, cortical representations for basic auditory attributes such as pitch and loudness are still poorly understood (Cariani and Micheyl, 2012), and in lieu of strong coding hypotheses, it is difficult either to confirm or entirely rule out prospective candidate codes. If complex temporal codes exist in central circuits that subserve the representation of all simple sensory attributes associated with events (e.g. the timing, duration, loudness, pitch, and timbre of a single musical note), they may involve longer delays with coarser temporal precisions or precise timing patterns that are difficult to observe because they are distributed across neurons.

These difficulties notwithstanding, temporal patterns of neuronal response associated with the discrete sensory and motor events themselves are very prominent in the cerebral cortex. Central temporal patterns related to temporal sequences of salient sensory events, such as musical rhythms and flashing lights, that occur on relatively coarse timescales (< 10/sec) are widely observed in evoked electrical and magnetic activity at many recording sites (Thatcher and John, 1977; John, 1967). Temporal pattern correlates of musical rhythmic expectancies are also observed (Zanto *et al.*, 2006; Fujioka *et al.*, 2012; Nozaradan, 2014). This evidence argues for temporal coding of musical rhythm at the cortical level. Stimulus-driven temporal response patterns encode the beats that are heard,

whereas stimulus-induced temporal patterns encode beats that are imagined or expected but not present acoustically and not heard. Coarse rhythmic patterns (<10 Hz) are supramodal, with widely distributed neural temporal responses that may explain how musical rhythm can provide a cross-modal temporal scaffold for movement and memory (Thaut, 2005). It should be noted that these stimulus-driven and stimulus-induced temporal patterns are distinct from the more widely studied endogenous brain rhythms that reflect neuronal population dynamics (Buzsáki, 2006). The relation between these neural dynamics and informational functions such as neural coding and grouping processes (Bullock, 1997; John, 1967; Giraud and Poeppel, 2012) remains an open question to this day. One possibility is that neuronal oscillations set the durations of temporal processing windows for interpreting latency-based temporal codes.

There also exist many classical examples of neuronal “assimilation of rhythms” in which neural assemblies internalize the temporal structure of their inputs during conditioning and then once internalized, produce those temporal patterns when activated (John, 1967; Morrell, 1967). Conditioning studies (Miller and Barnet, 1993) and anticipatory neural temporal prediction mechanisms (Schultz *et al.*, 1997) suggest that timelines of all events that lead up to reward are assimilated. If drive states (unmet goals) during conditioning are also broadcast as characteristic temporal patterns mixed in and assimilated with those of other internal events, then the resulting temporal memory trace can contain a representation of the nature and timing of goals pursued as well as specific event sequences and the rewards or punishments that ensued. Such a direct temporal tape-recorder-like representation, activated by similar sets of circumstances and encoded or read-out in faster-than-real-time can make prediction about the occurrence, timing, and likely hedonic consequences of future events.

Temporal processing architectures

We posit a universal temporal coding framework that can potentially handle all kinds of informational distinctions in the nervous system, matched universal short- and long-term memory mechanisms that can store and retrieve such distinctions (John, 1967), and a relatively small number of canonical neurocomputations that operate on temporal patterns. The basic mechanisms for coding, memory, and temporal processing are likely to be phylogenetically ancient. Most brain theories assume short-term reverberatory (echoic, working) memory coupled with long-term fixation. Temporal patterns could be regenerated by more local reverberatory loops over short time spans, re-broadcast by the hippocampus over intermediate spans, and encoded in molecular and synaptic connections over much longer terms. A universal coding system coupled with a universal memory system could provide a common language of distinctions for the brain that would parallel the general-purpose role of the genetic code in the cell.

Neural coding assumptions heavily influence our conceptions regarding the nature of neural information processing architectures. Channel codes lead directly to switchboard-like connectionist architectures that regard neurons as spike count (firing rate) integrators. In contrast, time codes lead to correlation theories of neural signal processing that are based on neurons as temporal coincidence detectors.

Temporal processing architectures take the form of neural time-delay networks (Licklider, 1959; MacKay, 1962), oscillatory networks (Greene, 1962; Buzsáki, 2006), axonal pulse computation trees (Pratt, 1989), synfire chains (Abeles, 2003), polychronous networks (Izhikevich, 2006), wave-interference networks (Heinz, 2010), and timing nets (Cariani, 2001a; Thatcher and John, 1977; John, 1967; Longuet-Higgins, 1989). Many

decades ago Karl Lashley envisioned wavelike, neuronal interference patterns as holograph-like, nonlocal alternatives to switchboard-type theories (Lashley, 1951; Orbach, 1998), and his successors proposed spatial and/or temporal mass-statistical mechanisms for implementing memory storage systems (Freeman, 1975; Pribram, 1971; Thatcher and John, 1977; John, 1967).

As a general rule, neurons in the central nervous system are organized into loops: brains are, first and foremost, “nets with circles” (McCulloch and Pitts, 1943) that have combinatorially large numbers of possible recurrent multisynaptic pathways, each with its own specific delay time. As a rule, regions of cerebral cortex are connected to their corresponding thalamic areas and to other (neighbouring and distant) cortical regions by reciprocal sets of projections that form re-entrant pathways. Another set of recurrent delay-paths can be found in the hippocampal formation, which is often regarded as a spatial, autoassociative memory. In this theory, temporal patterns of spikes circulate through neuronal loops, with the informational contents of the loops being related to the particular temporal patterns that are present. Through the senses and through internal pattern-generators, temporal patterns can be impressed on parts of the system and actively regenerated. The active delay loops that propagate spike patterns function as complex pattern-oscillators. The temporal pattern signals interact in local and global loops, mutually reinforcing or interfering with each other, eventually settling into stable, steady informational, dispositional states that determine and guide action.

Neural timing nets and temporal pattern-resonances

Neural timing nets are information processing architectures that operate entirely on temporal pulse patterns (Cariani, 2001a, 2002, 2004). A neural timing net model assumes representations and operations that stay in the time domain. Whereas connectionist architectures are “connection-centric” (all informational function depends on particular synaptic connection-weightings), neural timing networks are “signal-centric” (action lies in interactions between signals).

Both inputs and outputs of timing nets are temporal patterns of pulses that are processed through arrays of delay lines and pulse coincidence detectors. Feed-forward timing nets (FFTNs) consist of delay-coincidence arrays where input lines cross at many relative delays (Figure 1(b), left). Recurrent timing nets (RTNs) consist of coincidence arrays with delay loops that build up repeating temporal patterns within the loops (such as those produced by rhythmic musical stimuli). FFTNs implement cross-correlation-like operations, RTNs autocorrelation-like ones. Delays can be either monosynaptic (recurrent collaterals) or polysynaptic (delay-paths through networks), with synapses being either fixed or spike timing dependent (inhibited/facilitated by recent spike correlation history). Timing-dependent synapses support competition between signals and winner-take-all dynamics (facilitation of one set of signals inhibits others), as well as the possibility of regenerative, self-facilitating “synfire cycles”.

In timing nets, signals directly interact with each other, by selecting and mutually amplifying sub-patterns they have in common (“pattern-resonances”). A visual metaphor is shown in Figure 1(c). By acting as temporal pattern sieves, they can extract particular sub-patterns from mixtures, enabling signal combinations to be demultiplexed into their primitive components, operations critical for representational compositionality. Some of these basic operations, as applied to auditory features, are presented in (Cariani, 2001a, 2002).

The temporal signal-processing supports vector representations and operations that can potentially subserve Gestaltist grouping processes (Cariani, 2001a, 2004) and as well

as concept-network dynamics (Widdows, 2004). If we regard the constituent temporal patterns in the two spike trains as independent vector dimensions, then FFTNs multiply the two pattern vectors and output only those patterns that are common to the two inputs. The result is that the common sub-patterns in reciprocally connected local FFTNs resonate with each other, causing their multiplicative amplification. Ensembles of signals in different local regions (A, B, C, D, E and B, D, F, G) reinforce only those pattern-dimensions (B, D) that they share in common (Figure 1(b), right). Subsets of such mutually interacting signals will build up those feature primitives that their members have in common and this becomes a basis for perceptual grouping and conceptual coherence.

Timing networks also support content-addressable memory, by permitting high-dimensional representation spaces (Kanerva, 1988) to be searched in parallel. If a temporal probe pattern is presented to a reciprocally connected FFTN with fixed synapses, then that pattern will return if and only if the probe pattern is present in the inputs to the target FFTN. If correlated inputs can facilitate synapses, then injection of a temporal pattern into a timing network leads to its successive recurrent amplification and build up. If the facilitations extend beyond individual synapses to whole neurons, then other delay-paths are also facilitated. These kinds of dynamics can support priming of all neural assemblies that produce the injected pattern, which in turn permits content-addressable search. For example, an injected signal related to a particular attribute (“green”) or object (“elephant”) or goal (“how to satisfy thirst?”) resonates with all other neural assemblies that share the attribute or its components, amplifying their signal-sets, thereby “bringing them to mind”.

The signal primitives activated include associated motor programs for action that have followed goal presentations in the past, such that a search for addressing some particular goal prepares the system to act to achieve that goal. By directly including goal- and action-signals in neural representations, when a goal becomes urgent and its characteristic signal is injected into the rest of the system, pattern-resonance with the goal signal brings forth possible goal-directed actions. This kind of pragmatic, goal-directed content-addressable indexing is evident in animal and human minds, but is almost completely absent from digital computer search systems (which tend to index attributes of objects rather than their manifold possible uses).

The signal productions of a given neural assembly can be regarded as activation of a concept node, with pattern-resonances causing spreading activation (priming) of related concept nodes. Those nodes that have the most signals in common are activated first. Those neural assemblies with similar temporal patterns (encoded features) will mutually stabilize each other and suppress the signal productions of dissimilar patterns.

In cognitive science, feature and vector-based models of concept networks have been developed for lexical semantics (Schreuder and Flores D’arcais, 1989; Widdows, 2004). Many of these exhibit spreading activation through networks on the basis of similarity metrics and stabilization of mutually consistent meanings. There are also tantalizing analogies with Pask’s Conversation Theory, which describes how networks of concepts may be stabilized through mutual resonances and constructive processes. Pask’s theory of concept dynamics provides a “second-order” cybernetic explanation for how the concept-systems of actor-participants can evolve through convergent communicative interactions (Pask, 1996).

Although his theory is rather abstract and formal, it is evident that Pask had concrete, physical and neural implementations of concepts in his mind. His paper on his self-organizing electrochemical device was, after all, entitled “Physical analogues to the

growth of a concept” (Pask, 1959; Cariani, 1993). In other papers, one sees glimpses of how Pask (1980) thought his concept networks might be realized in brains and analog devices, using coupled oscillatory elements (pp. 371-375).

Working along similar lines in the 1960’s, Silvio Ceccato and his colleagues at the Italian Operational School used pulse oscillators to develop a mechanistic signal-processing cybernetic model of concept-based semantic operations (Von Glasersfeld, 2007). Ceccato regarded semantic compositions by observer-actors as concrete constructivist acts in and of themselves and considered their many second-order cybernetic epistemological implications (Parini, 2011). The current research programme in this intellectual lineage strives to identify primitive operations of cognitive information processing (“mind operational semantics”) and to develop a neutrally grounded implementations of these operations (“operational architectonics”) (Benedetti *et al.*, 2010). In a similar spirit, the theory proposed here seeks to explain concept-dynamics in terms of interacting temporal pattern signals that in effect carry out correlation operations on high-dimensional feature vectors.

There are other parallels with Pask’s concept networks as well. In the 1980’s, Paul Pangaro developed a digital computer programme in LISP named THOUGHTSTICKER to implement concept networks and their mechanics (Pangaro, 2001). The interactive programme, which preceded the advent of the World Wide Web and operated under very limited computational constraints, implemented cross-indexed hyperlinks into all other nearby, relevant nodes within a high-dimensional feature and concept space, much the same way that in this timing net theory spreading trajectories of pattern-resonances are conceived to activate successive neural assemblies and their signal productions (concept nodes).

Lastly, some account is needed of how new concepts (categories, perceptual symbols) might be created. Concept-formation is important for constructivist theories of mind and for modelling communication processes in which interacting minds coevolve their own meanings. New neural assemblies can potentially be formed by changing synaptic weights associated with sets of time delays such that a new temporal patterns is emitted when a particular combination of input features and goal signals are presented. The new, emergent temporal pattern would function as a “tag” that a particular combination of features with specific meaning to the system is present (Cariani, 2012). Such “perceptual symbols” (Barsalou, 1999) thus indicate a categorical distinction or concept that goes beyond an arbitrary combination of features.

It is conceivable that the aforementioned neuronal assimilation of rhythms is related to the formation of a new neural assembly. The neural assembly might generate this new tag from the composite pattern of the characteristic features combined with a training, reward signal. The operation should be reversible: by presenting the tag pattern, one should be able to weakly evoke the signals that generated it. A time-domain correlation mechanism for convolution and deconvolution operations that could fulfil this role has been proposed (Longuet-Higgins, 1989). The operation allows for the recovery of constituent features from concept-tags that would be generated via convolution.

Brains can be regarded as powerful, self-organizing content- and purpose-addressable search engines. Questions such as “what color is an elephant?” can be answered by presenting the concept-tag-signal that connotes elephant to cortical regions holding visual memories, thereby evoking imaginary images that weakly produce signals related to the colour grey which in turn activate the lexical node “grey” and the motor assemblies needed to utter the answer). Presentation of the tag-signal to auditory areas weakly

evokes their sounds, to olfactory areas, their smells. By generating a higher level tag pattern and presenting it to a local neural assembly, then seemingly abstract concepts can produce particulars. Through the dynamics of these temporal signal interactions, temporal codes, and timing nets may be able to realize the entailment meshes and concept-logics that are needed for thought and language.

Neurophenomenology

1227

Finally, in addition to accounting for informational functions that subserve behaviour, a full theory of brain function should also seek to explain the structure of subjective experiences that are produced by these systems. Such a theory should provide bridge laws that predict states of conscious awareness and their specific experiential contents (neurophenomenology). The best current hypotheses for the neural correlates of awareness involve recurrent patterned neural activity in global neuronal workspaces that subserve working memory (Dehaene, 2014). Those sets of signals that are actively regenerated in the global loops determine the experiential contents of conscious awareness, whereas the ability itself to coherently regenerate neural signals is held to be the organizational requisite for waking consciousness. The unitary nature of awareness is attributed to multimodal integration of neural information in a “global workspace” (Baars, 1988). Alternate states of consciousness (sleep, hypnosis, meditation, trance) involve different dynamical modes of local and/or global regeneration.

Current hypotheses about the neuronal basis of conscious awareness differ in their (usually implicit) neural coding assumptions. This is the question of how the specific contents of conscious awareness are encoded, i.e. precisely what aspects of neuronal activity patterns are regenerated. Connectionist theories assume that particular subsets of neurons are differentially activated via elevated or depressed firing rates, such that it which ensembles are active determines the contents of subjective experience (Rose, 2006). In this neural timing theory, the signals are circulating temporal patterns, such that circumstances that disrupt the coherent global regeneration of temporal patterns should abolish awareness entirely (anaesthesia, coma, seizure). In both kinds of theories awareness can be disrupted by suppressing (silencing) particular neuronal populations or by rendering incoherent neuronal patterns of activity (scrambling the messages) such that neuronal signals cannot be effectively regenerated (Cariani, 2000). In the temporal theory, neuronal spiking that is not structured such that can be interpreted as part of a code is effectively ignored by the system. In this respect, because it depends on the coherence of neuronal activity, the temporal theory is perhaps closest to E.R. John’s (2002) “hyperneuron” concept of correlated neural activity.

Organizational closure is a fundamental requisite for consciousness in a number of theories. In contrast to hypotheses that posit neural states must attain a threshold informational complexity for awareness (Tononi, 2005), the timing net model posits that the ability of the system to maintain steady informational states through coherent signal regeneration is key (i.e. closing an organizational loop rather than requisite complexity). Organizational closure is also a central requirement for other theories that rely on synchronized mass-oscillatory activity (Varela *et al.*, 2001; Thompson and Varela, 2001; Rudrauf *et al.*, 2003), re-entrant signalling (Edelman, 1992), thalamocortical self-selection of inputs (Llinas and Ribary, 1994), loops anchored in parietal first-person self-space maps (Pollen, 2011), or formal closure of process (Pask, 1979, 1981).

Organizational closure leads to informational autonomy and to the self. Those circular-causal processes generated from within are experienced as self-generated

thoughts, whereas those contingent sensory inputs that switch neural activity patterns from without are experienced as external sensations. Those autonomous sensorimotor routines that come to effectively control some aspect of the outer world, thereby eliminating contingency, become unconscious automatisms. The functional and phenomenal boundaries for the self, which inform our perceptions of agency in others and of our own sense of our selves, may flow directly from these domains of internal vs external causation (what our brains reliably control and what they do not control).

Conclusions

The timing net theory outlined here provides a general cybernetic, neural model of information processing in brains that utilizes temporal pattern pulse codes and temporal pulse coincidence computational operations. The timing theory provides an alternative to connectionist models of brain function that are based on average firing rates and specific interneural connectivities (synaptic weights). A universal, complex temporal pulse pattern coding scheme is proposed that permits multiplexing of neural signals in pulse trains of individual elements and the pulse-pattern statistics of ensembles of elements. A neural architecture consisting of tapped delay lines, delay elements, and pulse coincidence detectors with spike timing dependent plasticity operates on the complex temporal pulse patterns.

Reciprocally connected ensembles of these delay-coincidence arrays build up those temporal patterns that are common to both ensembles (pattern-resonance). Those sets of temporal pattern signal productions that are mutually reinforcing build up, circulate, and prevail in the network of loops formed from interconnected ensembles (an autopoiesis of neuronal signals). The set of signals circulating in the network at a given time constitute the functional state of the system and predispose (prime) the system for subsequent incoming neuronal signals (attention, context) and actions. The forms of the circulating signals that have been actively regenerated by the network beyond a threshold signal-to-noise ratio constitute the contents of conscious awareness at any moment (neurophenomenological bridge laws).

The theory as it stands is in a highly tentative, formative state. Further elaboration, refinement, and eventual empirical testing of several key elements is needed: universal temporal coding schemes (neural coding), pattern-regenerative processes (organizational stability), perceptual symbol (concept) formation, and reversible compositional operations (combinatorial meaning).

The neural coding problem for this theory entails formulation of a universal framework within which all internal distinctions can be expressed. Universal temporal coding schemes permit distinctions related to different informational functions (e.g. sensory attributes, cognitive representations, goal states) to be processed using the same analysis, storage, and retrieval operations. If all internal distinctions can be expressed in terms of complex pulse patterns (e.g. complex spike latency patterns), then one canonical set of temporal neurocomputational operations can suffice to realize all informational functions.

The theory holds that the functional state of the brain (i.e. a mental state) is determined by which sets of temporally coded signals are being actively regenerated in neural circuits at any given time. Exactly how such temporal patterns might be actively regenerated in networks with realistic neuronal parameters is an open problem. Its solution would entail demonstration of how spike pattern statistics of neuronal ensembles could be maintained over time, and thus how short-term temporal pattern memories could operate.

Temporal operations for perceptual symbol formation are needed to produce emergent tag patterns that signify grouped features and meaningful associations. The operation should be partially reversible: the symbolic tag pattern should be able to weakly elicit production of the lower level features that were originally associated. Reversible compositional operations are needed for combinatorial representational systems (concept combinations, language). Ideally perceptual symbol formation and compositional operations should be realizable through signal interactions (mechanics of high-dimensional temporal pattern vectors).

Despite its heterodox neural coding assumptions, the theory offers a heuristically useful neurocomputational alternative to connectionist theory that is flexible, purposive, informationally open, organizationally closed, and self-organizing.

References

- Abeles, M. (2003), "Synfire chains", in Arbib, M.A. (Ed.), *The Handbook of Brain Theory and Neural Networks*, 2nd ed., MIT Press, Cambridge, MA, pp. 1143-1146.
- Arbib, M.A. (1972), *The Metaphorical Brain: An Introduction to Cybernetics as Artificial Intelligence and Brain Theory*, Wiley-Interscience, New York, NY.
- Baars, B.J. (1988), *A Cognitive Theory of Consciousness*, Cambridge University Press, Cambridge.
- Barrett, L. and Bliss-Moreau, E. (2009), "Affect as a psychological primitive", *Advances in Experimental Social Psychology*, Vol. 41 No. 41, pp. 167-218.
- Barsalou, L.W. (1999), "Perceptual symbol systems", *Behavioral and Brain Sciences*, Vol. 22, pp. 577-660.
- Benedetti, G., Marchetti, G., Fingelhurts, A.A. and Fingelhurts, A.A. (2010), "Mind operational semantics and brain operational architectonics: a putative correspondence", *The Open Imaging Journal*, Vol. 4, pp. 53-69.
- Boring, E.G. (1942), *Sensation and Perception in the History of Experimental Psychology*, Appleton-Century-Crofts, New York, NY.
- Bullock, T.H. (1997), "Signals and signs in the nervous system: the dynamic anatomy of electrical activity is probably information-rich", *Proc Natl Acad Sci USA*, Vol. 94 No. 1, pp. 1-6.
- Buzsáki, G. (2006), *Rhythms of the Brain*, Oxford University Press, Oxford and New York, NY.
- Cariani, P. (1993), "To evolve an ear: epistemological implications of Gordon Pask's electrochemical devices", *Systems Research*, Vol. 10 No. 3, pp. 19-33.
- Cariani, P. (1995), "As if time really mattered: temporal strategies for neural coding of sensory information", *Communication and Cognition – Artificial Intelligence (CC-AI)*, Vol. 12 Nos 1-2, pp. 161-229 (Reprinted in K. Pribram (Ed.) (1994), *Origins: Brain and Self-Organization*, Lawrence Erlbaum, Hillsdale, NJ, pp. 208-252).
- Cariani, P. (1999), "Temporal coding of periodicity pitch in the auditory system: an overview", *Neural Plasticity*, Vol. 6 No. 4, pp. 147-172.
- Cariani, P. (2000), "Regenerative process in life and mind", in Chandler, J.L.R. and Van de Vijver, G. (Eds), *Closure: Emergent Organizations and Their Dynamics*, Vol. 901, Annals of the New York Academy of Sciences, New York, NY, pp. 26-34.
- Cariani, P. (2001a), "Neural timing nets", *Neural Networks*, Vol. 14 Nos 6-7, pp. 737-753.
- Cariani, P. (2001b), "Temporal coding of sensory information in the brain", *Acoust. Sci. & Tech*, Vol. 22 No. 2, pp. 77-84.
- Cariani, P. (2002), "Temporal codes, timing nets, and music perception", *J. New Music Res*, Vol. 30 No. 2, pp. 107-136.

-
- Cariani, P. (2011), "The semiotics of cybernetic percept-action systems", *International Journal of Signs and Semiotic Systems*, Vol. 1 No. 1, pp. 1-17.
- Cariani, P. (2012), "Creating new primitives in minds and machines", in McCormack, J. and D'invorno, M. (Eds), *Computers and Creativity*, Springer, New York, NY, pp. 395-430.
- Cariani, P. and Micheyl, C. (2012), "Towards a theory of information processing in the auditory cortex", in Poeppel, D., Overath, T. and Popper, A. (Eds), *Human Auditory Cortex: Springer Handbook of Auditory Research*, Springer, New York, NY, pp. 351-390.
- Cariani, P.A. (2004), "Temporal codes and computations for sensory representation and scene analysis", *IEEE Transactions on Neural Networks*, Vol. 15 No. 5, pp. 1100-1111.
- Cariani, P.A. and 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*, Vol. 76 No. 3, pp. 1698-1734.
- Dehaene, S. (2014), *Consciousness and The Brain: Deciphering How The Brain Codes Our Thoughts*, Viking, New York, NY.
- Edelman, G.M. (1992), *Bright Air, Brilliant Fire: On the Matter of the Mind*, BasicBooks, New York, NY.
- Emmers, R. (1981), *Pain: A Spike-Interval Coded Message in the Brain*, Raven Press, New York, NY.
- Freeman, W.J. (1975), *Mass Action in the Nervous System*, Academic Press, New York, NY.
- Fujioka, T., Trainor, L.J., Large, E.W. and Ross, B. (2012), "Internalized timing of isochronous sounds is represented in neuromagnetic beta oscillations", *J. Neurosci*, Vol. 32 No. 5, pp. 1791-1802.
- Giraud, A.L. and Poeppel, D. (2012), "Cortical oscillations and speech processing: emerging computational principles and operations", *Nature Neuroscience*, Vol. 15 No. 4, pp. 511-517.
- Greene, P.H. (1962), "On looking for neural networks and 'cell assemblies' that underlie behavior. I. Mathematical model. II. Neural realization of a mathematical model", *Bull. Math. Biophys*, Vol. 24, September, pp. 247-275, 395-411.
- Heinz, G. (2010), "Introduction to wave interference networks", in Kyamakya, K., Unger, H. and Halang, W.A. (Eds), *Mallorca Workshop 2010 – Autonomous Systems (Smart Systems Technologies)*, Camp de Mar, Mallorca, Spain, Shaker-Verlag GmbH, Aachen, pp. 1-9.
- Izhikevich, E.M. (2006), "Polychronization: computation with spikes", *Neural Comput*, Vol. 18 No. 2, pp. 245-282.
- John, E.R. (1967), *Mechanisms of Memory*, Wiley, New York, NY.
- John, E.R. (2002), "The neurophysics of consciousness", *Brain Research Reviews*, Vol. 39 No. 1, pp. 1-28.
- Kanerva, P. (1988), *Sparse Distributed Memory*, MIT Press, Cambridge, MA.
- Kilmer, W. and McCulloch, W.S. (1969), "The reticular formation command and control system", in Leibovic, K.N. (Ed.), *Information Processing in the Nervous System*, Springer-Verlag, New York, NY, pp. 297-307.
- Lashley, K.S. (1951), "The problem of serial order in behavior", in Jeffress, L.A. (Ed.), *Cerebral Mechanisms of Behavior*, Wiley, New York, NY, pp. 112-146.
- Licklider, J.C.R. (1959), "Three auditory theories", in Koch, S. (Ed.), *Psychology: A Study of a Science. Study I. Conceptual and Systematic*, McGraw-Hill, New York, NY, pp. 41-144.
- Llinas, R.R. and Ribary, U. (1994), "Perception as an oneiric-like state modulated by the senses", in Koch, C. and Davis, J.L. (Eds), *Large-Scale Neuronal Theories of the Brain*, MIT Press, Cambridge, MA, pp. 111-124.
- Longuet-Higgins, H.C. (1989), "A mechanism for the storage of temporal correlations", in Durbin, R., Miall, C. and Mitchison, G. (Eds), *The Computing Neuron*, Addison-Wesley, Wokingham.

-
- McCulloch, W.S. and Pitts, W.H. (1943), "A logical calculus of the ideas immanent in nervous activity", in McCulloch, W.S. (Ed.), *Embodiments of Mind (1965)*, MIT Press, Cambridge, MA, pp. 19-39.
- MacKay, D.M. (1962), "Self-organization in the time domain", in Yovitts, M.C., Jacobi, G.T. and Goldstein, G.D. (Eds), *Self-Organizing Systems 1962*, Spartan Books, Washington, DC, pp. 37-48.
- Miller, R.R. and Barnet, R.C. (1993), "The role of time in elementary associations", *Current Directions in Psychological Science*, Vol. 2 No. 4, pp. 106-111.
- Morrell, F. (1967), "Electrical signs of sensory coding", in Quarten, G.C., Melnechuck, T. and Schmitt, F.O. (Eds), *The Neurosciences: A Study Program*, Rockefeller University Press, New York, NY, pp. 452-469.
- Nozaradan, S. (2014), "Exploring how musical rhythm entrains brain activity with electroencephalogram frequency-tagging", *Philos Trans R Soc Lond B Biol Sci*, Vol. 369 No. 1658, p. 20130393. doi: 10.1098/rstb.2013.0393.
- Orbach, J. (1998), *The Neuropsychological Theories of Lashley and Hebb*, University Press of America, Lanham, MD.
- Pangaro, P. (2001), "Thoughtsticker 1986: a personal history of conversation theory in software, and its progenitor, Gordon Pask", *Kybernetes*, Vol. 30 Nos 5/6, pp. 790-807.
- Parini, P. (2011), "Ernst von Glasersfeld and the Italian operational school", *Constructivist Foundations*, Vol. 6 No. 2, pp. 140-149.
- Pask, G. (1959), *Physical Analogues to the Growth of a Concept. Mechanization of Thought Processes*, Vol II, H.M.S.O, London.
- Pask, G. (1979), "Consciousness", *Journal of Cybernetics*, Vol. 9 No. 3, pp. 211-258.
- Pask, G. (1980), "Developments in conversation theory – part 1", *Int. J. Man-Machine Studies*, Vol. 13, pp. 357-411.
- Pask, G. (1981), "Organizational closure of potentially conscious systems", in Zeleny, M. (Ed.), *Autopoiesis: A Theory of Living Organization*, North Holland, New York, NY, pp. 265-308.
- Pask, G. (1996), "Heinz von Foester's self organization, the progenitor of conversation and interaction theories", *Systems Research*, Vol. 13 No. 3, pp. 349-362.
- Pollen, D.A. (2011), "On the emergence of primary visual perception", *Cereb Cortex*, Vol. 21 No. 9, pp. 1941-1953.
- Powers, W. (1973), *Behavior: The Control of Perception*, Aldine, New York, NY.
- Pratt, G. (1989), "Pulse computation", PhD thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA.
- Pribram, K.H. (1971), *Languages of the Brain: Experimental Paradoxes and Principles in Neurophysiology*, Prentice-Hall, New York, NY.
- Rieke, F., Warland, D., De Ruyter, R., Steveninck, V. and Bialek, W. (1997), *Spikes: Exploring the Neural Code*, MIT Press, Cambridge, MA.
- Rose, D. (2006), *Consciousness: Philosophical, Psychological and Neural Theories*, Oxford University Press, Oxford and New York, NY.
- Rudrauf, D., Lutz, A., Cosmelli, D., Lachaux, J.P. and Le Van Quyen, M. (2003), "From autopoiesis to neurophenomenology: Francisco Varela's exploration of the biophysics of being", *Biol Res*, Vol. 36 No. 1, pp. 27-65.
- Schreuder, R. and Flores D'arcais, G.B. (1989), "Psycholinguistic issues in the lexical representation of meaning", in Marslen-Wilson, W. (Ed.), *Lexical Representation and Process*, MIT Press, Cambridge, MA, pp. 409-446.

-
- Schultz, W., Dayan, P. and Montague, P.R. (1997), "A neural substrate of prediction and reward", *Science*, Vol. 275 No. 5306, pp. 1593-1599.
- Sommerhoff, G. (1974), *Logic of the Living Brain*, John Wiley, London.
- Swanson, L.W. (2012), *Brain Architecture: Understanding The Basic Plan*, Oxford University Press, New York, NY.
- Thatcher, R.W. and John, E.R. (1977), *Functional Neuroscience, Vol. I. Foundations of Cognitive Processes*, Lawrence Erlbaum, Hillsdale, NJ.
- Thaut, M. (2005), *Rhythm, Music, and The Brain: Scientific Foundations and Clinical Applications*, Routledge, New York, NY.
- Thompson, E. and Varela, F.J. (2001), "Radical embodiment: neural dynamics and consciousness", *Trends Cogn Sci*, Vol. 5 No. 10, pp. 418-425.
- Tononi, G. (2005), "Consciousness, information integration, and the brain", *Prog Brain Res*, Vol. 150, pp. 109-126.
- Uttal, W.R. (Ed.) (1972), *Sensory Coding: Selected Readings*, Little-Brown, Boston, MA.
- Uttal, W.R. (1973), *The Psychobiology of Sensory Coding*, Harper and Row, New York, NY.
- Varela, F., Lachaux, J.P., Rodriguez, E. and Martinerie, J. (2001), "The brainweb: phase synchronization and large-scale integration", *Nat Rev Neurosci*, Vol. 2 No. 4, pp. 229-239.
- Verschure, P.F., Pennartz, C.M. and Pezzulo, G. (2014), "The why, what, where, when and how of goal-directed choice: neuronal and computational principles", *Philos Trans R Soc Lond B Biol Sci*, Vol. 369 No. 1655.
- Von Glasersfeld, E. (2007), "Cybernetics and the theory of knowledge", in Larochelle, M. (Ed.), *Key Works in Radical Constructivism*, Sense Publishers, Rotterdam, pp. 153-169.
- Widdows, D. (2004), *Geometry and Meaning*, CSLI Publications, Stanford, CA.
- Zanto, T.P., Snyder, J.S. and Large, E.W. (2006), "Neural correlates of rhythmic expectancy", *Advances in Cognitive Psychology*, Vol. 2 Nos 2-3, pp. 221-231.

About the author

Dr Peter Cariani (BS 1978, biology MIT, 1978, MS1982, PhD 1989 Systems Science Binghamton University; available at: www.cariani.com) has worked on a wide range of scientific and philosophical problems in theoretical biology, cybernetics, auditory neuroscience, and music perception. Dr Cariani is currently a Senior Research Scientist in the Hearing Research Center at the Boston University and a Clinical Instructor in Otology and Laryngology at the Harvard Medical School. He teaches courses related to the neuropsychology of music at the Harvard and the MIT. Dr Peter Cariani can be contacted at: cariani@bu.edu

permission.

Mihai Nadin *Editor*

Anticipation and Medicine



Springer

Temporal Memory Traces as Anticipatory Mechanisms

Peter Cariani

Abstract Brains can be considered as goal-seeking correlation systems that use past experience to predict future events so as to guide appropriate behavior. Brains can also be considered as neural signal processing systems that utilize temporal codes, neural timing architectures operating on them, and time-domain, tape-recorder-like memory mechanisms that store and recall temporal spike patterns. If temporal memory traces can also be read out in faster-than-real-time, then these can serve as an advisory mechanism to guide prospective behavior by simulating the neural signals generated from time courses of past events, actions, and the respective hedonic consequences that previously occurred under similar circumstances. Short-term memory stores based on active regeneration of neuronal signals in networks of delay paths could subservise short-term temporal expectancies based on recent history. Polymer-based molecular mechanisms that map time-to-polymer chain position and vice versa could provide vehicles for storing and reading out permanent, long-term memory traces.

Keywords Neural timing nets • Neural codes • Pitch • Rhythm • Auditory scene analysis • Temporal codes • Expectancy • Music perception • Engram

1 Introduction

As the aphorism goes, “the purpose of remembering the past is to predict the future.” Anticipation involves both predicting future situations and events and preparing for them. Anticipation not only projects *what* will occur but also *when* and *where* it will occur, as well as what to do about it. Anticipatory mechanisms enable organisms to use past experience to act in a manner appropriate for future conditions. This chapter proposes novel anticipatory neural memory mechanisms that are based on neural time codes and temporal pattern memory traces.

P. Cariani (✉)
Hearing Research Center, Boston University, Boston, USA
e-mail: cariani@bu.edu

© Springer International Publishing Switzerland 2017
M. Nadin (ed.), *Anticipation and Medicine*, DOI 10.1007/978-3-319-45142-8_7

105

2 Anticipatory Systems

“Anticipation” is a general notion. The theoretical biologist Robert Rosen coined the specific term “anticipatory system” [1–3] and defined it as a system in which prospective future states determine present behaviors [4]. The papers in this volume are the product of the last of three conferences inspired by Rosen’s and Nadin’s ideas about anticipatory systems [5, 6].

2.1 *Purposive Systems as Functional Organizations*

Anticipatory systems are those systems that have embedded goal-states: they are organized in such a way that their action realizes desired future states. The idea has much in common with naturalistically grounded teleologies (“teleonomies”) of purposive systems. In the east, Russian physiologists and psychologists (Anokhin, Sudakov, Bernstein, et al.) developed general theories of functional systems [7–9]. In the west, notions of purposive, feedback control systems formed the basis of the early cybernetics movement [10–14].

Rosen and his mentor, Nicolas Rashevsky, were strong proponents of a non-reductionist, relational, theoretical biology that focused on questions of *organization* as explanations of functions, rather than appeal to mechanistic reductionism [15, 16]. Rosen offered the parable of an amoeba in a pot of water, before and after boiling. The live amoeba and the dead one share the same molecular constituents, but the organization of the system has been altered by boiling in such a way that the amoeba was no longer able to regenerate its parts and its organization. Rosen mounted deep criticisms of the machine and computer metaphors for describing living systems, not because the parts in some way violate the laws of physics, but because, in describing living systems solely in terms of trajectories of parts, one misses the organizational relations that make the system a persistent, coherent entity. Knowledge of parts is useful certainly for designing drugs, but it does alone not tell us how to go about building stable, regenerative organisms [17]. For that, as Rashevsky and Rosen foresaw, one needs a theory of mutually stabilizing relations.¹

¹Such biological system theories have deep implications for medicine. Much of our current understanding of disease in terms of “molecular medicine” is grounded in linear chains of interactions between molecular parts. Many therapies simplistically attempt to control one variable (e.g., blood sugar concentrations) using one or two interventions (insulin) without considering the circular-causal nature of networks of metabolic loops that can stymie such interventions. Only if we are able to model the whole set of systemic interactions and relations can we anticipate what the system will do in the short and long term. Once we have an adequate systems theory of biological organization, we will gain the deeper understanding needed for how to design therapeutic interventions that have self-sustaining effects such that the need for further interventions becomes self-limiting.

2.2 *Anticipatory Systems in Animals with Nervous Systems*

Although the most obvious examples of anticipation involve animal behaviors mediated by learning and memory, many non-neural examples of biological anticipatory capabilities abound. For example, in many plants and animals, developmental stages are orchestrated to occur at favorable seasons of the year (e.g., seed germination in early spring). The mechanisms for favorable timing of developmental stages arise in environments with strong seasonal variations. The environmental variations create positive selective pressures for anticipation, such that those lineages, whose individual time development enhances survival and reproduction, will tend to persist longer than those that don't. Because of the cyclical, predictable nature of seasonal changes, timing strategies that worked better in previous cycles will continue to work better in present and future cycles. Thus anticipatory timing mechanisms appropriate for coping with the future can evolve, provided that similar situations recur.

Animals are motile organisms that cannot produce their own food. In contrast to fungi, which absorb nutrients, animals ingest and digest their food. As a consequence, most animals must move to find food, such that the immediate environments within which they must orient and transport themselves are ever changing. Animal lineages evolved nervous systems that coordinate the actions of effector organs contingent on the sensed states of immediate surrounds and on current internal goal-related states. Embedded goal-related states include the needs of the organism for survival (e.g., satisfaction of system-goals of homeostasis, self-repair, growth), and reproduction. Those lineages of organisms that evolved more effective embedded goal mechanisms for survival and reproduction tended to persist. In choosing actions contingent on percepts and active goals, organisms in effect anticipate which actions will be most appropriate in satisfying those goals.

On evolutionary timescales, variation, construction, and selection processes yield organisms that are better adapted in their particular ecological contexts for more reliable (survival and) reproduction. During the lifespans of organisms with nervous systems, neural learning processes shape percept-action mappings contingent on past experience and reward. So even in the most primitive kind of adaptive percept-action systems, there is anticipation in the sense that the results of previous experiences and successful performances continually modify system structure and behavior to guide future action.

In both evolution and learning cases, memory mechanisms encode the past and make it available for anticipation of what actions can be most appropriate in the current state. In the evolutionary case, the memory lies (mostly) in the genetic sequences that, shaped by selective pressures and construction constraints, persisted in the lineage. In the learning case, the memory lies in short-term memory traces that guide behavior based on the immediate past and present, and in more permanent long-term memory traces that can guide behavior that is based on the deeper past.

2.3 Organization of Perception and Action

Animals with nervous systems can be characterized in terms of purposive, percept-action systems. They have sensory receptors that permit them to make distinctions on their surrounds; effector organs (mainly muscles) that permit them to influence their environs (action); and nervous systems that permit coordination of action contingent on behavior. Aside from sleep or other dormant states, there is a constant, ongoing cycle of percepts, coordinations, actions, and subsequent environmental changes (Fig. 1, bold arrows).

In parallel with percept-action cycles are internal cyclical neuronal dynamics that steer behavior from moment to moment and over the long run (Fig. 2). These include the neuronal dynamics of competing internal goals, internal modal system-states (e.g., waking/sleeping, affective states), cognitive and deliberative processes, attention, action-selection, and the influence of long-term memories.

Embedded in all nervous systems of animals are feedback-driven goal mechanisms that steer behavior in a manner that reliably satisfy basic organismic imperatives of survival and reproduction (e.g., maintain oxygen/water homeostasis, find food sources, avoid predators, find mates). Competitive dynamics of current goals determine which goals are paramount at any given moment such that their

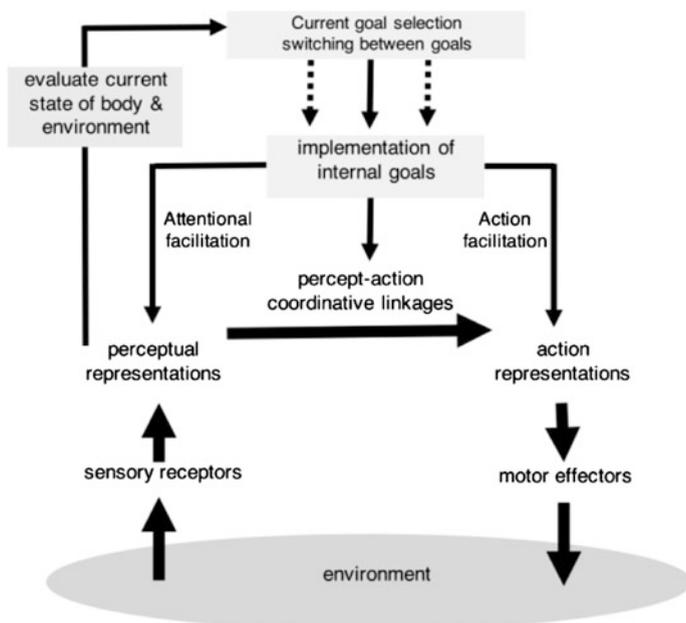


Fig. 1 Percept-coordination-action cycles and goal-directed steering of percept-action coordinations

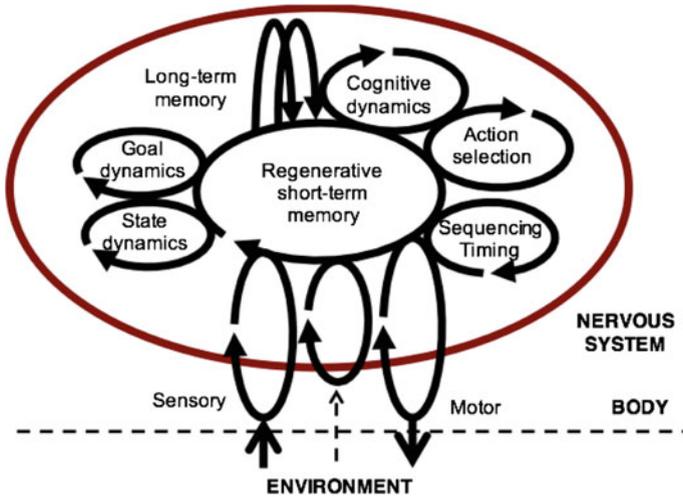


Fig. 2 Functional organization of informational dynamics in brains

drive states partially determine behavioral choices (action selections).² This is basically “dominance principle” of the Russian psychologist Ukhtomsky (1845–1942) [9]:

... in the nervous system, there is at each moment only one active dominating dynamic structure or constellation of excitation, which is associated with the most actual, urgent ongoing needs and desires. This excitation structure plays the role of a situational nervous network, an agency for organizing the physiological and behavioral response directed to satisfying these needs. At the same time all other goals and desires are suppressed” [18]. What is the difference between a bodily physiological mechanism in an animal and a technical mechanism? Firstly, the former is generated during the course of the reaction itself. Secondly, once chosen, the behavior of a technical mechanism is secured once and for all by its construction, whereas in a reflex apparatus it’s possible to successively realize as many different mechanisms as there are available degrees of freedom in the system. Each of the successively realized mechanisms is achieved through the active inhibition of all available processes, except one. (Ukhtomsky, quoted in [18])

The notion of competitive goal dynamics fits extremely well into current conceptions of brains as sets of competing circuits that steer behavior, for better or worse, towards particular sets of goals (“implementation of internal goals” in Fig. 1). The notion of the inhibitory suppression of non-dominant goals fits very well with the emerging conception of the basal ganglia/striatum as a generalized double-inhibition braking system in which brakes are selectively released to

²Drive states only partially determine actions, because arguably, actions taken also depend upon which action-alternatives (affordances) are perceived by the organism to be immediately available. Those goals that have obvious, apparent means of attainment may be more attractive for action. Perception of options for effective action can feed back to change which drive states become dominant.

facilitate task-relevant thalamic sensory and motor channels. The resting state is a state of inhibition of sensory and motor systems in which neural loop gains are slightly attenuating. The release of inhibition changes these gains to slightly amplifying, thereby facilitating neuronal sensory and motor signals related to the current dominant goal.

In addition to competitive goal dynamics, virtually all animals switch their global system-states between discrete modes of operation and behavior, (e.g., waking-sleep-hibernation cycles and behavioral modes such as eating, excreting, hunting, fleeing, communicating, exploring, mating) [19]. Affective states, which can be regarded as internal assessments of the current overall “state of the organism” also modulate behavior choices by increasing propensities for different modes of action (e.g., fight/flee/approach). Behavior is jointly determined by internal states (goal-drive state, system-mode, affective states), as well as by the perceived current state of the environment and the perceived goal-satisfaction action-possibilities that the situation affords (Fig. 1). The situation perceived by the animal is in turn jointly determined by the state of the environment and the animal’s perceptual systems.

2.4 General Types of Functional Organizations

In my view, such animals can be categorized in terms of the functionalities that their organizations afford. Animals are *living systems* because they actively regenerate their organization (material components and relations). This is the core idea underlying conceptions of self-production systems [20], autopoietic systems [21, 22], metabolism-repair systems [23, 24], self-reproducing robots [25], autocatalytic nets [26], self-modifying systems [27], and semantically-closed self-interpreting construction systems [28]. They are *semiotic systems* because their internal operation and resulting behavior can be described in terms of sign distinctions conveyed via neural codes [29, 30]; and they are *autonomous purposive systems* because they are mainly driven by internal goals.

They have their own *agency* to the extent that they have embedded goals (internal motivation), requisite ability (the right mechanisms needed for action), and sufficient freedom of action (autonomy) to reliably achieve particular goals. They are *anticipatory systems* if they have learning and memory mechanisms that allow them to project the past into the present so as to evaluate future consequences of current courses of action. If one defines these different attributes in terms of these specific kinds of material organizations, then a system, such as an autonomous robot, need not be living to be semiotic, autonomous, purposive, or anticipatory, or to exhibit agency. Underlying these different types of functional organization is the Aristotelian notion of *hylomorphism*.

2.5 *Hylomorphism as an Ontology for Functional Organization*

Hylomorphism is an ontology of functional organizations embedded in matter. Aristotle adopted the hylomorphic framework in formulating his theories of life and mind [31–33].

Life, purpose, meaning, and even conscious awareness are properties of material systems that are organized in particular ways. Explanations based on organization (the system is organized so as to realize a particular goal, i.e., it has a “final cause”) are complementary to reductionist, causal explanations based solely on physical properties of parts. Purposive, goal-directed systems are material systems that are organized so as to realize particular goal end-states that in effect become their final causes.

An example of a simple purposive system is a thermostatically regulated heating system that is organized so as to maintain the temperature of a room within a particular range. The “final cause” of the system is the end-state target temperature range that is determined by the thermostat. (Because of its organization and material realization, the system seeks the corresponding temperature states, and, provided that the system is working properly, the thermostat setting “predicts” the final temperature state of the system.)

Hylomorphism is a functionalist ontology to the extent that functional organization can be abstracted from particular material substrates. One can design a thermostatically regulated heating system in terms of functions of and relations between components without specifying exactly how thermostat control mechanisms and heating/cooling elements are to be realized materially (and different material implementations can realize comparable behavioral functions). However, unlike platonic ontologies based entirely on ideal forms, a hylomorphic ontology is materially grounded. In order to realize functions within the material world, organizations must be realized in *some* material form. It is not enough to replicate form; the organization must be fleshed out, embodied, such that it interacts with and changes other parts of the material world.

3 Anticipation and Memory

Memory is a process that entails the maintenance of a distinction through time, and thus it is a semiotic process that is invariant with respect to time. Anticipatory prediction involves estimating the course of future events based on the (remembered) past and present (Fig. 3).

Nervous systems evolved to coordinate behavior. Coordination without memory is possible where mappings between percepts and actions do not change with experience. However, once these coordinative mappings can be modified on the basis of experience, then the effects of past experiences can carry over into present

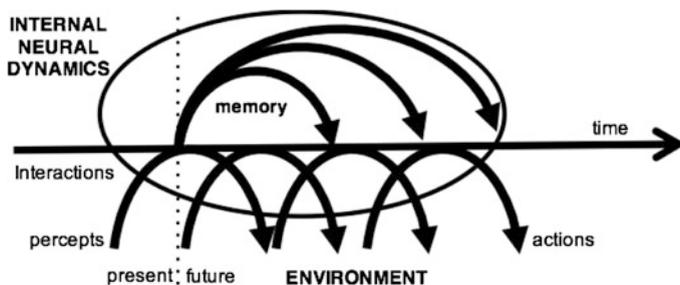


Fig. 3 Memory enables present distinctions to carry over to influence future actions

and future. This kind of simple adaptive modification of behavior does not require explicit storage and retrieval processes.

In biological organisms and nervous systems, anticipation involves not only *what* situations and events are expected, but also *when* and *where* they are expected to occur. In animals with nervous systems, anticipation involves (usually implicit) understanding of the contingent structure of the world (modeling) for deliberative purposes: to decide whether positive action needs to be taken, to determine what actions are available (perceived affordances), and if so, what action is most likely to satisfy those system-goals that are currently of highest priority (goal satisfaction).

3.1 Short- and Long-Term Memory

Standard theories of memory posit a labile, short-term memory coupled with a permanent long-term memory (Fig. 4). There is large literature, old and new, in psychology on the characteristics and nature of memory [34–36]. Many treatments further subdivide different types of memory by modality, the nature of the items stored, temporal processing windows, and temporal persistence, while others seek universal frameworks.

Short-term memory, broadly construed, provides a temporary store of neural signals related to current and recent perceptions, thoughts, affects, motivations, as well items maintained via working memory or recalled from long-term memory. Since the 1930s, the neural mechanisms that subserve short-term memory have been conceived explicitly in terms of neuronal reverberatory processes, i.e., neural activity patterns that are actively maintained and self-sustaining. The sustained firing of neurons that permits these activity patterns to persist is facilitated by activation of N-methyl-D-aspartate (NMDA) receptors that create the biophysical conditions for long-term potentiation (LTP) and spike-timing-dependent plasticity (STDP).

How the specific contents of these temporary memories are coded in neuronal activity patterns is the neural coding problem as it applies to memory. If the

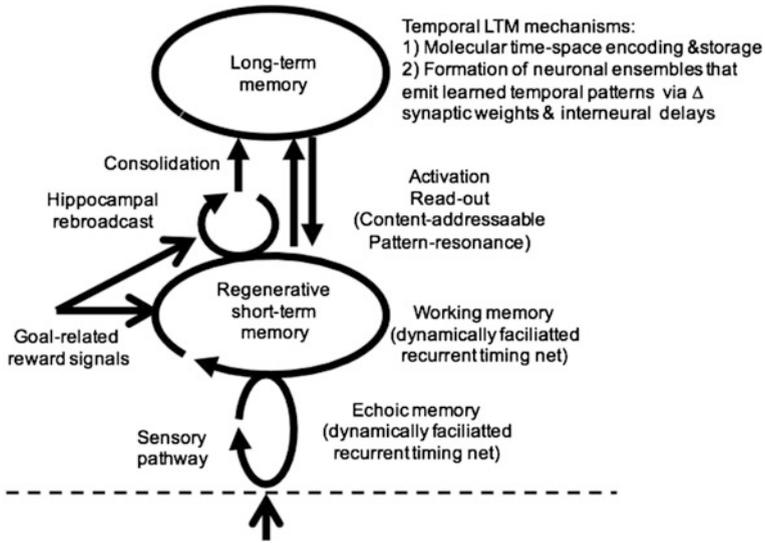


Fig. 4 Short-term regenerative memory and long-term memory. Neuronal signals related to current goal-states are maintained in regenerative short-term working memory and eventually consolidated into long-term memory traces. In this proposed scheme, the contents of short-term memory stores consist of complex temporal spike patterns that are actively regenerated in correlation-facilitated delay paths. Long-term temporal memory traces are activated by corresponding temporal patterns present in working memory, enabling pattern-resonance and content-addressability

information is encoded such that specific subsets of neurons have sustained, elevated firing rates, then the reverberating patterns likely involve persistent activation of these specific neuron subsets. If information is encoded in temporal spike patterns, as is proposed here, then the reverberating patterns likely involve persistent regeneration of spike patterns within neuronal circuits.

The hippocampal formation is the bridge between short-term and long-term memory stores. It appears to be responsible both for maintaining neural activity patterns in short-term memory and for replaying neuronal patterns associated with significant events such that they can be consolidated into long-term memory, mainly during sleep. This work originally came out of animal maze-running experiments, where “place cells,” which encode distinct maze locations, were observed to fire in the temporal sequences of the maze running. Recently, “time cells,” which encode the timings of events and places, have been found in the hippocampus and elsewhere [37]. During states of sleep or during periods of awake reflection, stored event-sequences can be replayed at faster-than-real-time, enabling them to function as predictive reward mechanisms [38].

Long-term memory is permanent: it can survive sleep, seizure, general anesthesia, and long periods of coma. Once formed, some types of long-term memories can last the entire lifetime of an individual.

3.2 *Music, Memory, and Anticipation*

Music perception offers a rich set of examples of the actions of short- and long-term memory mechanisms whose operations span timescales from seconds to lifetimes [39, 40]. Because music is a universally familiar medium that involves sequences of discrete events that unfold in time within the evolving context of the remembered recent present, it provides an excellent springboard for studying the role of time in mind and brain.

In music, every note-event creates an expectancy of the next note-event such that sequences of notes create pattern expectancies of succeeding notes. Repeating a sequence of events immediately groups the events in the sequence into a coherent “chunk” and creates a strong expectancy that the pattern sequence will repeat yet again. The pattern expectancies can involve temporal patterns of the note timings (rhythmic patterns), note accent patterns (meter), pitch sequences (motifs, melodies), and timbral sequences. Temporal grouping processes, which were extensively investigated by the Gestaltists, play an essential role in shaping these expectancies [39, 41, 42].

The brain is extremely good at detecting patterns of events and attributes that recur, and it can do this over many timescales.³ In music, periodic sound patterns whose waveforms rapidly repeat (25–4000 Hz) produce musical pitches (that can carry a melody), whereas slower periodic patterns of sonic events (10 Hz or less), be they musical notes, clangs, or speech fragments, create rhythmic expectancies. Every repeating sound creates a strong expectancy of its continuation, and violations of this expectancy are highly noticeable. The repeating pattern is grouped into a unitary whole (a “chunk”). If the pattern is a fast-repeating acoustic waveform with a repetition rate greater than about 25 Hz, we perceive a pitch at the repetition rate. The tonal quality (timbre) of the sustained sound is determined by the form of the repeated waveform (or in Fourier terms, its spectral distribution and shape).

In music, repeating temporal patterns of onsets and offsets of discrete events with rates of less than roughly 10 per second are perceived as rhythms. Different temporal patterns are heard as different rhythms. Here the repeating events are grouped together into a chunk (“groove”), and after a few repetitions, expectations that each of the events will recur at a given time in the sequence and with the same attributes rapidly develop. This implies that the representation of the rhythmic pattern contains not only information about the sequence of events and their respective attributes (e.g., pitch, timbre, loudness, duration, location), but also a timeline of events. The memory trace of a rhythmic pattern essentially replicates the timeline of the events (the timings of event onsets, mainly, but also offsets that co-encode event-durations).

³Oliver Selfridge once related to me his experience with early computers with mechanical relays. The programmers would set up the computer and would stay in the room periods of time while the computations ran. After a while they would begin to notice long and elaborate repeating irregular sound patterns that were being produced by the relays as the program executed long, iterated complex instruction loops.

What is the nature of this memory trace? Although it has been conventionally assumed that temporal relations are converted to spatial patterns of neuronal activation, there have been a number of suggestions in the past that the memory traces themselves might be temporal patterns. In his book on the psychology of time, Fraisse [43] references the “cyclochronism” theory of the Russian psychologist Popov. Longuet-Higgins had proposed various nonlocal, holographic memory schemes based on frequency-domain Fourier decompositions (holophone model, [44]). Although these have the merit of using the oscillatory dynamics of neural populations, they had very limited storage capacity. Much more promising was a later time-domain mechanism based on temporal correlations between spikes [45]. Roy John proposed that evoked, induced, and triggered temporal patterns of response could subservise general mechanisms for memory [34, 46, 47].

We propose that temporal patterns of spikes might provide a direct basis for temporal memory trace mechanisms. In essence, temporally patterned stimuli impress their temporal structure on the time structure of neural spikes, such that the temporal patterns of spikes can serve as an iconic time-domain representation of the stimulus.

A great deal of evidence in the auditory system points towards temporal codes for pitch and rhythm, albeit at different levels of the system. The most obvious neural correlates of musical pitch lie in spike timing patterns—distributions of interspike intervals—in the auditory nerve, brainstem, and midbrain [48, 49]. Those for rhythm can be found at all levels of the system, from auditory nerve to auditory cortex and beyond. The onsets and offsets of every note produce well-timed spikes in large numbers of cortical neurons [50], such that temporal patterns of note-events replicate the temporal structures of their rhythms [51, 52]. These evoked neuronal temporal responses provide the necessary inputs for various oscillatory mechanisms of rhythmic and metrical expectancy that have been proposed [53]. (Possible direct time codes for rhythm and their relationship to oscillatory dynamics are discussed a bit more below in Sect. 5.2.2.)

3.3 *Mismatch Negativity and Short-Term Expectancies*

Short-term musical memory is very sensitive to the timing of events; and there are neural responses that are widely observed in event potentials that appear to be related to short-term temporal expectancies and their violations. Event potentials are averaged electrical or magnetic signals triggered (aligned in time) by the onset of a particular event, such as a note-event or even the change in a note-event pattern.

A so-called “mismatch negativity response” (MMN) is observed when an event is presented repetitively (“standard”) and then some change is made in a subsequent event (“deviant”). Such MMN-like responses are widely used to study the dynamics of musical expectancies and their violations [50, 54, 55].

If the event is a musical note, any perceptible change in the physical attributes of the note, such as a change in sound level (loudness), periodicity (pitch), spectrum

(timbre), attack (timbre), duration, or location (apparent location), will produce a neural response that differs from the response to the standard event. Here the context of the standard event pattern has created an expectancy that is then violated by the deviant event. The mismatch negativity is computed by subtracting the time-pattern of the averaged response to deviant events (triggered on their onsets) from that of the standards. Depending on the nature of the change in events, the mismatch negativity peaks at a characteristic latency after the beginning of the deviation from the standard, i.e., at the onset of the deviant stimulus. This latency is typically on the order of 100-200 ms for changes in basic auditory perceptual attributes.

The MMN is evoked more or less automatically. It does not require subjects to attend to the stimuli, and indeed can be observed in sleeping subjects, infants, and many animals. MMN responses to a standard metrical musical pattern followed by the same (syncopated) pattern in which one of the beats is omitted are observed in newborn infants. MMN-like responses are observed (sometimes reported under a different name) in many different species, sensory modalities, and brain regions.

MMN-like responses can also be seen for “higher level” patterns and attributes. Complex rhythmic and/or melodic patterns of note-events can be presented as standards, and a pattern that deviates in some respect (e.g., a change in the periodicity (pitch) of one of the notes) will create an MMN closely following the time point of the deviation. MMN-like responses with longer latencies are observed for syntactic and semantic violations and yet other kinds of more abstract attributes.

MMN responses are sensitive not only to changes in the perceptual and cognitive attributes of discrete events, but also to their timings. If a regular metrical sequence is set up as the expected standard pattern, then deviations in the timing of subsequent events (leads or lags relative to the expected timing of the beat) will evoke MMN responses, again with a latency that depends on the timing of the expectancy-violating mismatch. In music, these expectancy violations form the basis for expressive timing, intentional manipulation of the timings of notes to convey and evoke emotions.

MMN phenomena suggest the existence of canonical neural temporal comparison mechanisms. It is as if a timeline of events is being built up, maintained in short-term memory, and compared with incoming temporal patterns. In this chapter, I propose a complex neural time code in which both the attributes of events and their relative timings are encoded and simple neuronal delay-and-compare mechanisms (recurrent timing nets) that would produce similar kinds of behaviors. What is needed is a mechanism that both builds up an expectancy of patterns of events when they recur, and computes the deviation of the incoming stream of new events from what was expected at that particular time point.

There is an ongoing debate about the nature and meaning of the MMN [56]. Some current theories of the MMN hold that a memory trace is formed when the standard is repeated and that there is a comparison of incoming neural activity patterns against this memory trace that was constructed from very recent experience. Others hold that the memory trace itself may be embedded in the responses of

ensembles of cortical neurons. Still others reject the notion of organized memory traces in favor of explaining the MMN in terms of neuronal adaptation processes.

It is notable that early studies of electrical conditioning in single neurons, conducted before MMN was discovered, found that cortical neurons assimilated rhythmic patterns (10 Hz flashes of light) that were presented in conjunction with correlated electrical pulses similar in many respects to a reward signal [34, 57, 58]. The stimuli were presented over and over, paired with electrical pulses; and over tens of trials the temporal response pattern of the neurons came to resemble that of the stimulus (10 Hz firing pattern). When the stimulus was then abruptly changed to 1 Hz, the slower flashes evoked the 10 Hz pattern for many repetitions, but eventually the assimilated rhythm was extinguished.

3.4 Temporal Theories of Associative Memory

Predictive timing is a key element of anticipatory behavior. It is often important to know when a reward will come. The relative timing of rewards and the events that lead up to them has been an ongoing concern of theories of learning.

It has also been observed that spike timings of dopaminergic neurons reflect discrepancies between anticipated and observed courses of the neural concomitants of events associated with rewards [59, 60]. This discovery has spawned a host of adaptive temporal prediction models.

Many studies in animal and human conditioning suggest that the timings of all correlated events relative to the arrival of a reward are implicitly and intrinsically stored in both short and long term memory, such that any of them can serve as anticipatory temporal predictors [61]. This temporal coding of memory hypothesis thus proposes that “the temporal conditions (e.g. the CS-US interstimulus interval) are not merely catalysts in the formation of associations, but are also a part of the content of learning” [62]. The hypothesis further asserts that animals can build temporal maps from relationships between events that were never physically paired, “that is, temporal information from different training situations which have a common element can be integrated based on super-positioning of the common element in different temporal maps” [63]. This means that systematic maps of temporal relations between events can be built up from separate experiences of subsets of events. Such maps of temporal relations can then subserve anticipatory prediction—each event becomes a predictor for other temporally correlated events [64].

4 A Temporal Theory of Brain Function

Brains implement anticipatory predictions that subsequently guide behavior. In this paper we propose a high-level theory of brain function based on temporal pulse pattern codes that can be actively regenerated, stored, and retrieved. In this theory,

prediction and steering of behavior are achieved by encoding and retrieving temporal patterns of spikes associated with internal events connected to percepts, actions, rewards, and punishments. Although it shares many assumptions with mainstream connectionist theory, (e.g., recurrent connections, distributed coding and processing), this proposed theory differs from connectionism in its fundamental neural coding assumptions. Whereas connectionism is based entirely on channel-activation codes (which neurons fire at which average rates), the alternative neural architectures envisioned here rely on temporal codes, i.e., neural codes that encode distinctions in temporal patterns of spikes.

A general theory of brain function requires specification of several basic aspects.

1. *Neural codes.* What are the system's signals? A neural coding scheme based on spike timing patterns must be capable of representing all the distinctions that we make (e.g., encoding all of the attributes of objects, events, and their relations and their compositions).
2. *Neural networks.* What processing architectures are needed to realize the informational signal processing operations that the system performs? A neural processing architecture capable of operating on temporal patterns in order to carry out informational operations—such as detections, discriminations, pattern recognitions, invariances, transformations, and groupings—is required.
3. *Memory mechanisms.* How are informational distinctions encoded and decoded in memory? Here mechanisms of short- and long-term memory that can store and recall temporally coded temporal patterns in a content-addressable manner are needed.

These different aspects of the system need to be compatible with each other. The nature of the neural codes that bear informational distinctions heavily determines the nature of neural signal processing architectures and the memory mechanisms needed to utilize them. Conversely, the available neural mechanisms for processing, storing, and retrieving information heavily constrain what kinds of codes the system can use.

5 A Universal Coding Framework Based on Complex Temporal Spike Patterns

5.1 *The Neural Coding Problem*

Understanding the nature of the neural code (the “neural coding problem”) in central circuits is arguably the most fundamental problem facing neuroscience today. Without an understanding of the precise nature of the “signals of the system,” we cannot have a firm grasp of the nature of information processing and storage in brains. Neuroscience today is in a situation comparable to cellular biology and genetics before DNA nucleotide sequences came to be understood as

the primary vehicle for inheriting and expressing genetic information. However, neural coding is rarely explicitly mentioned as an unsolved problem in neuroscience. Although interest in neural coding has undergone cyclical changes from decade to decade, it has never yet risen to the forefront of neuroscience.

A common tacit assumption in mainstream neuroscience is that the coding problem has already been solved, that the brain is a large, complex connectionist network. The instantaneous functional states of the system are thought to be patterns of average firing rates across neurons, and its structural informational states are thought to be characterizable in terms of interneural connection weights. The instantaneous functional state (mental state), taken together with the structural state (the “connectome”), is thought to determine how the system behaves. Note that these assumptions only hold to the extent that spike timing and temporal relations between neurons (intra- and inter-neuronal time delays) have no significant role in informational functions. Solving the neural coding problem is critical for interpreting the functional significance of specific neuronal connection patterns.

Neural codes, as discussed here, are the functional signals of the system, those neuronal activity patterns that have functional significance for information processing in the brain. Different kinds of informational distinctions (e.g., perceptual attributes, thoughts, desires, affective states) and their specific alternatives (e.g., for the visual attribute of color, the distinctions of red vs. blue vs. green vs. yellow) are mediated through specific patterns of neural spiking activity. These specific types of patterns and different patterns within a type constitute the neural codes [65–69].⁴

This notion of coding is related to the conception of a sign in semiotics as a distinction that has functional significance for its user. A sign, in Gregory Bateson’s phrase, is a “difference that make a difference.” A neural code is a pattern of activity that makes a functional difference in the brain, i.e. a difference that alters internal functional states and subsequent behavior. To go further, some spike patterns may be meaningful, having consequences for internal states or overt behaviors, whereas others that are not interpretable within a coding scheme may not constitute a coherent internal message, and so would not be meaningful to the system. Examples of the latter might include uncorrelated “spontaneous” spike patterns, spikes generated during epileptic seizures, and incoherent firing patterns produced under general anesthesia.

5.2 *Types of Neural Codes: Channel Codes and Temporal Codes*

Neural pulse codes can be divided into two types: channel codes and temporal codes. In channel codes, activation of particular subsets of neurons (channels)

⁴The functional definition of neural coding is different from formal, information-theoretic Shannonian estimates of channel capacities that are independent of whether or not the system makes use of the different states.

conveys distinctions (e.g., through across-neuron firing rate profiles), whereas in temporal codes particular temporal patterns of spikes convey distinctions. Temporal codes can be further divided into those codes that depend on temporal patterns of spikes irrespective of their times-of-arrival (spike latency) and those codes that depend on the relative latencies of spikes. Temporal pattern codes based on inter-spike intervals appear to subservise auditory pitch and cutaneous flutter-vibration sensations, whereas relative time-of-arrival codes appear to subservise various stimulus localization mechanisms based on temporal cross-correlation in audition, somatoception, and electroception.

Connectionist theory adopts the assumption that the central codes operant in brains are firing-rate channel codes, whereas the neural timing net theory outlined here posits that these are complex temporal pattern codes. Thus, for most of the history of modern neuroscience, channel coding has been adopted as the conventional, default assumption and with it, connectionism as the default neural network assumption. However, a significant minority opinion has involved proposal of various kinds of temporal codes as alternatives [65]. Early examples include Rutherford's "telephone theory of neural coding," Troland's temporal modulation theory of hearing, Wever's temporal volley theory, the Jeffress' model for binaural localization, and Licklider's duplex model for pitch perception. There has likewise always been an alternative tradition for temporal processing architectures as well [70–74]. We have discussed many of these various neural coding schemes elsewhere [69, 75, 76].

5.2.1 Types of Temporal Codes

In simple temporal codes (Fig. 5a), one temporal parameter conveys one perceptual distinction. For example, in the early auditory system, times between spikes (interspike intervals) carry information about the periodicities of sounds. At these early auditory stations, the neural code for pitch involves the mass statistics of interspike intervals amongst whole populations of neurons [49, 69, 79]. The pitch that is heard corresponds to the most common interspike intervals that are produced by the population. The sensation of flutter-vibration has a similar basis in simple interspike interval patterns. Binaural auditory localization in the horizontal plane utilizes sub-millisecond spike timing differences produced by corresponding neurons in neural pathways that originate in the two ears. Analogous examples exist in nearly every sense modality [76].

Complex temporal pattern codes can be formed from combinations of simple temporal-pattern primitives (Fig. 5a–c). Here different orthogonal *types* of temporal patterns encode different independent primitive features. For example, for a musical note-event, the different dimensions of pitch, loudness, duration, location, and tonal qualities (timbral distinctions) would be conveyed via different types of temporal patterns of spikes that were produced concurrently within some population of auditory neurons (Fig. 5c). Combinations of specific temporal patterns present at any given time form a feature-vector whose dimensions are determined by the pattern types.

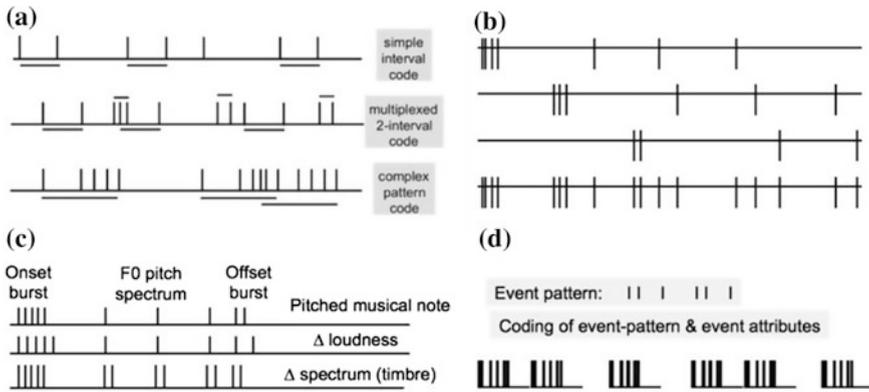


Fig. 5 Temporal codes. Idealized spike trains illustrate different coding schemes. **a** Simple and complex temporal pattern codes: *top* interspike interval code (e.g., for pitch or flutter-vibration); *middle* multiplexing of two different interspike intervals related to different types of information; *bottom* complex pulse pattern code. **b** Multiplexing of three complex temporal patterns associated with different types of cutaneous sensation (after [76]); *bottom* spike train shows interleaving of the different signals. **c** Hypothetical spike latency pattern scheme for encoding different attributes of auditory events (pitch, loudness, timbre, duration). **d** Hypothetical universal scheme for encoding event-patterns that includes event attributes and the relative timings of events

Complex temporal codes can potentially be conveyed in single units or in the pattern-statistics of spike time patterns in ensembles and populations. The various patterns can be interleaved or embedded in other patterns of spikes (Fig. 5b), permitting signal multiplexing (concurrent transmission of multiple types of signals over the same neuronal transmission lines). A complex temporal coding scheme that was proposed in the past for multiple cutaneous sensory qualities [78] provides a concrete example of how such codes might be organized.⁵

A neural code must convey two types of information: the type of distinction the signal conveys (e.g., pitch or color), and the attribute distinction itself (e.g., a particular pitch or color). In channel codes, the identity of the channel (which neuron, as determined by its place in the network, its interconnectivities) conveys informational type, whereas patterns of channel activations convey different attribute values. For channel-coding schemes, neural channel-identities maintained via specific interconnections are absolutely critical for function. A neural firing rate is meaningless to the rest of the system without the identity of the neuron that is firing. Connectionist networks are completely dependent on the maintenance of highly regulated connectivities (“synaptic weights”).

In contrast, because complex temporal codes embed the type of the signal in the form of the message, highly specific interneural connectivities and signal transmission paths are no longer essential for function. Temporal codes thus permit

⁵As far as we know, Emmers’ findings have not been replicated experimentally by others, so the neuronal reality of this powerful, multidimensional coding scheme is unproven.

“signals to be liberated from wires.” As long as the patterns can propagate to other populations, the specific paths they take does not change the nature of the messages sent. Thus this mode of neural communication enables broadcast modes of signal distribution. Neuronal assemblies downstream can be selectively tuned such that they respond to particular temporal patterns embedded in their inputs. Broadcast, multiplexing, and selective tuning enable decentralized communications in which tuned neural assemblies can respond only to those incoming signal patterns that are relevant to their functional roles. Here a “neural assembly” is used in the Hebb-Lashley sense of a functional organization of a set of neurons—a given neuron may participate in many different neural assemblies that are organized around different tasks [80]. Some codes may be restricted to local neuronal populations (e.g., restricted cortical regions) that handle specific types of information, whereas others may be propagated more globally, re-broadcast by the hippocampus, and then consolidated into long term memory.

Finally, complex temporal codes can represent patterns of events that occur over different time scales. If a sequence of musical notes is played, such as a motif or a melody, a temporal code that can encode all of the attributes associated with the individual events can also encode the timings of the events (Fig. 5e). Thus rhythmic pattern and musical meter can be encoded on coarser timescales and the same coding framework can handle different levels of musical organization.

5.2.2 Evidence for Temporal Codes

Temporal codes are found in a very wide range of sensory systems [66–69, 75–79]. Temporal codes in sensory systems have been found with sub-microsecond spike timing precisions (electroreception), microsecond precisions (bat and cetacean echolocation), sub-millisecond precisions (auditory, somatoception, vision), and still coarser (olfaction, gustation).

Generally speaking, auditory systems tend to have the highest frequencies of the synchronization of spikes with stimuli. In the auditory systems of barn owls, who use spike timing to localize their prey in the dark, spikes fire in time with the fine structure of sounds up to periodicities of 10 kHz. In humans and cats, primary auditory neurons phase lock up to roughly 5 kHz. Next, spike synchronizations to electrical shocks delivered to the skin approach rates up to about 1 kHz. Visual neurons lock to modulations in luminance up to roughly 100 Hz.

Temporal codes are found not only in sensory systems in which spikes follow the fine time structure of the stimulus, such as hearing and touch, but also in sensory systems such as vision, where eye movement transforms spatial luminance patterns into correlated temporal patterns at the retina. Even in the chemical senses of smell and taste, and in color vision, differences in the temporal response properties of sensory receptors produce corresponding characteristic temporal patterns of spiking.

In general, the most obvious temporal codes are found near sensory receptor surfaces. In early stations of the auditory system, spike timing is most abundant and

its functional role most obvious. However, as one proceeds up ascending sensory pathways, stimulus-related timing information is mixed with other kinds of information and in some cases smeared out such that stimulus-related fine timing information above roughly 100 Hz becomes successively less apparent as one proceeds to the cortical level [75]. In the auditory system, despite considerable progress, cortical representations for basic auditory attributes such as pitch and loudness are still poorly understood [81].

In lieu of strong coding hypotheses, it is difficult either to confirm or entirely rule out prospective candidate codes at the cortical level. Thus the complex temporal pattern-coding scheme outlined here is a very tentative hypothesis. If complex temporal codes do exist in central circuits that subserve the representation of all simple sensory attributes associated with events (e.g., the timing, duration, loudness, pitch, and timbre of a single musical note), they may involve timing patterns that are difficult to observe because they are distributed across neurons and/or not rigidly synchronized either with the stimulus or each other.

Although temporal patterns associated with fine temporal structure above a few hundred Hz are not abundant at the cortical level, temporal patterns of neuronal response associated with the slower successions of onsets and offsets of discrete sensory and motor events (periodicities <10 Hz) are very prominent. The precision of the coding of these onsets, on the order of 100–200 μ s, is maintained all the way up the auditory pathway. Temporal patterns related to patterns of musical events, flashing lights, electric shocks, and tactile pulses are widely observed in evoked electrical and magnetic activity over large parts of the cerebral cortex [34, 46].

Musical rhythm is thus a prime candidate for temporal coding. Patterns of event onsets are seen widely in averaged electrical and magnetic auditory evoked potentials and also more recently even in single trial stimulus presentations. At present, it is possible to determine which of two auditory streams a listener is attending on the basis of the corresponding rhythmic pattern of neuronal response in an EEG or MEG signal. Even beats that are expected, but not presented or heard, as in a syncopated meter, produce observable responses at the times when the acoustically sounded beat was expected to arrive [82–84]. Actively imagining a rhythmic pattern facilitates the pattern in neuronal cortical populations such that it can be observed in EEG recordings.

Rhythms in speech, though less regular and not as well defined as their counterparts in music, likewise produce corresponding temporal patterns that reflect acoustic contrasts [52, 85]. Recently rhythmic patterns of neuronal response associated with different levels of sentential organization (e.g., syllables, words, phrases, sentences) have been observed at the cortical level [86].

All of these phenomena argue for direct temporal coding of rhythm in music and speech at all levels of auditory processing. Coarse rhythmic patterns (<10 Hz) are supra-modal, with neural temporal responses that are very widely distributed across cortical regions [46]. The ubiquity of these correlated neural response patterns may explain how musical rhythm can provide a cross-modal temporal scaffold for movement and memory [87].

In the last decade, there has been a renewed interest in brain rhythms, and many of these phenomena have been interpreted in terms of oscillatory dynamics of neuronal populations [86, 87–89], rather than under the rubric of temporal codes [49, 51, 90]. It is important to distinguish between evoked, exogenous rhythms that reflect driving stimulus periodicities both from endogenous rhythms produced by intrinsic dynamics of neuronal excitation and recovery and from induced rhythms that are triggered or released by external stimulus events. The focus is usually on entrainment and induced rhythms rather than stimulus-locked synchronization or stimulus-related periodicities. Synchronization is often regarded as a passive process, in contrast to “active modification of ongoing brain activity” in entrainment [51]. Note, however, that for every observed oscillation, there is an associated temporal organization of spiking activity. Recent evidence points to substantial functional roles for both evoked and induced rhythms in the grouping and analysis of speech and musical events. Oscillatory dynamics of cortical neurons may govern temporal processing windows for music and speech [89, 91, 92] in a manner that limits the rate at which events and their various attributes can be accurately represented. In terms of neural coding, the durations of these windows may place constraints on how fast complex spike latency codes (such as the code of Fig. 5c), which require different readout times for different attributes, can be produced and processed.

6 Neural Architectures for Temporal Processing

What kinds of neural architectures would be needed to utilize a temporal coding framework such as the one outlined above?

6.1 *Basic Plan of the Brain*

The basic structural plan of animal brains [93] is well-conserved phylogenetically. Despite its apparent neuroanatomical complexity, brains consist of a relatively small number of component subsystems and neuro-computational architectures. First and foremost, as neuroscientists have understood for more than a century now, the brain is a network of recurrent pathways. These have variously been called loops, neural circuits, re-entrant paths, and nets with circles.

The brain can thus be regarded as a network of neuronal circuits, i.e., large numbers of interconnected neuronal loops that contain excitatory and inhibitory neurons with local connections, and excitatory (and sometimes inhibitory) neurons with longer range projections. Many different canonical neural circuits have this recurrent organization (e.g., thalamocortical loops, cortico-cortico re-entrant pathways, and hippocampal loops). The systematic sets of recurrent pathways in the hippocampus have been often regarded as computational substrates for an

auto-associative memory mechanism. Different regions of the cerebral cortex are reciprocally connected to neighboring ones by local connections and to more distant cortical regions by white-matter long-range axonal tracts. However, closed cyclical chains of excitatory neurons without inhibition create positive feedback loops that quickly saturate. By regulating the amount of inhibition, the loop-gains of these circuits can be modulated from their resting, slightly attenuating state, in which incoming neuronal activity patterns die out, to states of attention.

For the most part, neuroscience has been more focused on the neuroanatomy of connections, i.e., “the connectome,” than on temporal relations within and between the various neuronal populations. However, for timing theories of brain function, each connection between neural elements has not only a synaptic weight, but also time delays associated with axonal and synaptic transmission. Each of these various loops have characteristic time delays associated with them: short-delays for local circuits and longer ones amongst more distant neuronal ensembles. The more numerous unmyelinated axons, with their slow conduction velocities and long conduction times, yield much longer transmission delays than myelinated axons. In addition, there are delay processes that are inherent in the recovery dynamics of the neuronal elements that can perform many of the same neural signal processing functions as transmission delays.

6.2 Neural Time-Delay Architectures

Purely connectionist networks do not represent time explicitly, except as sequences of changing spike rates. Time-delay networks, on the other hand, include time delays between elements that allow them to interconvert temporal patterns and spatial activation patterns. Early time-delay neuro-computational architectures were proposed for that utilized binaural time disparities for localizing sounds and monaural interspike intervals for perceiving periodicity pitches (Jeffress and Licklider models, [52, 59]).

Time-delay networks use coincidence-detector elements with short time integration windows for handling temporal patterns of spikes, and rate-integrator elements with long time integration windows for converting spike coincidences to average firing rates. This “coincidence counting” allows them to interface with connectionist architectures. In effect, temporal codes are converted to rate-channel codes. Time-delay architectures can be flexible in their ability to handle both temporal and spatial information: by tuning delays, one can change synaptic efficacies, and vice-versa.

6.3 Neural Timing Network Architectures

After many years of searching for alternatives to both connectionist rate-place and time-delay neural network schemes, I proposed yet a third kind of neural network,

which I called *neural timing nets*. For the most part, early work on these networks involved demonstrating the various time-domain operations that could be elegantly carried out [94–96].

Temporal pattern codes allow simple form transformations (position shift invariance), and time warping of patterns yields tempo-invariance of rhythms, transposition invariance of pitch sequences, and magnification invariance of spatial forms. Using these kinds of temporal representations, separation of independently moving or changing forms can be effected. One can easily separate objects on the basis of invariant relational patterns of elements within objects (fusion, grouping) vs. the changing relations between elements of different, independently moving objects (separation).

Neural timing networks consist of arrays of delays and coincidence elements that operate on temporally-coded inputs and produce temporally coded outputs. Essentially, everything is kept in the time domain, and neural signals can interact with each other to sort out common temporal subpatterns. Whereas connectionist and most time-delay architectures are “connection-centric” (all informational function depends on particular synaptic connection weightings), neural timing networks are “signal-centric” (action lies in interactions between signals: “signal dynamics”).

The signal-centric nature of the networks (and networks processing based on signal dynamics) sidesteps many of the problems of connectionist and time-delay network architectures, in that precise and elaborate point-to-point connections are not needed for such networks to function. It is enough to bring the various neural signals into the same regions at approximately the same time. By operating on the temporal pattern statistics of ensembles of neurons, as long as there are some points of interaction, it no longer matters whether this or that neuron produced this or that output.

Feedforward timing nets (FFTNS) are arrays of coincidence detectors and delay lines that cause temporally patterned signals to interact. Various correlation and convolutional operations can be carried out, enabling multiplexing and demultiplexing of signal primitives. In FFTNS the spike train signals collide, interact, interfere, and/or mutually amplify each other, essentially performing correlation-like filtering signal processing operations in the time domain.

Compared to connectionist networks, the temporally coded representations and signal processing operations are more iconic and analog in character and more parallel in implementation. Template matching can be realized by injecting a temporal pattern archetype into the network, which will serve to amplify any incoming temporal pattern signals that have significant correlations with it. Content-addressable search can likewise be realized by injecting temporal patterns related to the features that one is interested in. Other neural signals circulating within the network will interact with the search signal if and only if they have feature-related temporal subpatterns in common. Essentially complex temporal pattern signals can implement a vectorial representation in which the signals themselves can sort out those dimensions that they have in common. The informational operations involve “pattern resonances” [97]. The processing scheme as it currently stands is provisional and still in a rudimentary state of development.

Nonetheless, it appears to be much more flexible than any connectionist scheme we have seen to date.

7 Temporal Mechanisms for Short-Term Regenerative Memory

Most high level accounts of brain function posit networks of recurrent pathways (re-entrant loops, neural circuits) that support a dynamic, working short-term memory coupled to a more permanent long-term memory mechanism that permits storage and retrieval of relevant patterns (Fig. 2). Neuronal activity patterns that have hedonic salience for the animal (i.e., are part of a string of events that leads to significant reward or punishment) are rebroadcast by the hippocampal formation such that the patterns are maintained in working memory and later consolidated and fixed in long-term memory. This rebroadcast can replicate event sequences at faster-than-real-time rates.

Conventionally, the nature of short-term memory is commonly assumed to involve subsets of specific neurons in recurrent neural circuits that maintain higher rates of activity, whereas long-term memory is thought to involve changes in the effective connectivities between neurons at synapses (in neural network terms, “synaptic weights”). Thus short-term memory is conceived in terms of a complex reverberation pattern of neuronal activations, while long-term memory is thought to entail more permanent synaptic changes.

In addition to feedforward timing nets, there also can be recurrent timing nets (RTNs), in which there are arrays of delay loops that span a wide range of recurrence times. RTNs were initially conceived as models for pitch- and rhythm-based grouping and separation mechanisms [94–96]. For both pitch and rhythm, repeating waveforms and temporal event patterns respectively create strong temporal expectancies and groupings. Our auditory systems easily separate concurrent sounds with different fundamental frequencies (F_0 s), such that we are able to hear out different musical instruments and voices when they are mixed together. RTNs, which act in a manner similar to adaptive comb filters, separate out the respective temporal patterns of multiple speakers with different voice pitches and of multiple musical instruments playing different notes.

Recurrent timing nets are perhaps the simplest kind of reverberating, temporal memory that can be imagined (Fig. 6a). Here an incoming pattern is compared with a delayed circulating pattern; and if the recurrence time of the delay loop is equal to the repetition time of a repeating pattern, then the pattern is facilitated (builds up) in that particular loop. If the difference between the circulating and the incoming pattern is also computed, then the difference signal can be fed into the array. Each delay loop creates an expectation of what the next incoming signal fragment will be. (The expectation is a primitive anticipation in that the recent past is used to predict the near future.) In the case of a repeating pattern of auditory events, each

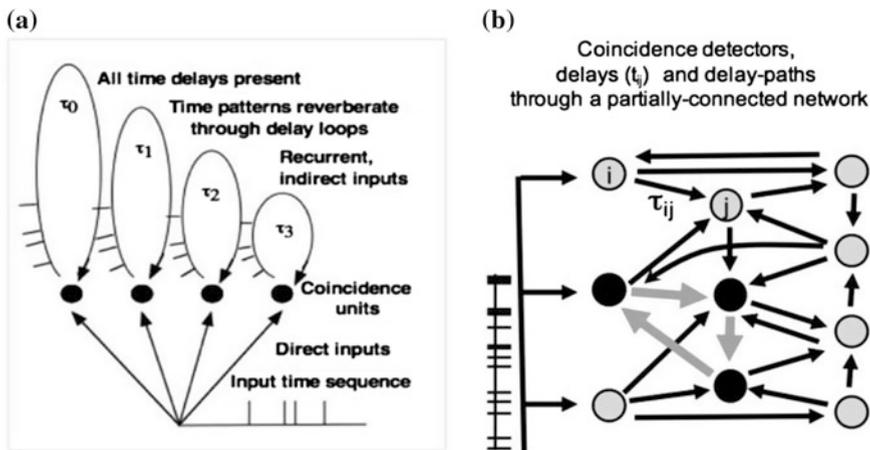


Fig. 6 Recurrent timing nets (RTNs). **a** An array of delay loops and correlation-facilitated coincidence detectors. A periodic pulse pattern fed into the network maximally facilitates itself in the delay loop, whose recurrence time is equal to the duration of the repeated pulse pattern, not unlike a time-domain implementation of a comb filter. Such a network stores temporal patterns in the delay loops. Multiple periodic patterns sort themselves out in the different delay loops, which then function as complex pulse pattern-oscillators. **b** Alternative implementation of a neural timing net in a richly interconnected network of coincidence-facilitating elements. A pulse pattern is fed into the leftmost, input layer and delay paths corresponding to the internal delay structure of the stimulus are facilitated (*gray arrows, black coincidence elements*). The reverberating patterns are regenerated within the network (*synfire cycles*)

delay loop is making a temporal prediction about when the next event will occur. And if the event-attributes (timing, loudness, pitch, timbre, duration, location) are also encoded in time via a complex, multidimensional time code, then it creates expectancies for those also.

The RTN delays can be either monosynaptic (recurrent collaterals) or polysynaptic (delay paths through networks, Fig. 6b). The number of recurrent monosynaptic 2-element paths in a fully connected network of N elements is on the order of N^2 . The number of paths increases combinatorially with maximum path-length M , roughly as M^N . Although the brain is thought to be more like a small world rather than a fully interconnected network [98], all neurons are thought to be interconnected by at most three or four interneurons. The number of delay paths is still astronomical, far greater than even the combinatorics of individual synapses (because each synapse connecting two neurons has a delay associated with it).

If the synapses are spike-timing dependent (inhibited/facilitated by recent spike correlation history), then the repeating pattern will flow through those delay-paths that have recurrence times equal to the pattern repetition time. Other paths with other delays that are not in the pattern will be temporarily inhibited. In this manner,

to whatever degree there is repetition, the recurring time structure of the rhythm input will build up in the network. If there are only locally repeating patterns, say ABCABCCABBCDEFAB, which has a maximum repeated pattern length of 2 (AB), then the network will revert to the probabilities of shorter sequences. This becomes a neural implementation of a variable-order Markov chain that can adapt to variable N-gram lengths. The system thus predicts specific, longer sequences when those have been presented in the not-too-distant past, but in lieu of repeating sequences reverts to Bayesian statistics.

Timing-dependent synapses support competition between signals and winner-take-all dynamics (facilitation of one set of signals inhibits others), as well as the possibility of regenerative, self-facilitating “synfire cycles.” These differ from synfire chains [73, 99, 100] in that the temporal pattern statistics, rather than which neurons are firing, encode particular attribute distinctions. The regenerative cycles in effect would constitute a temporal echoic and working memory buffer that would hold the temporal patterns (maintaining the statistics of the patterns) such that they could be compared with incoming ones.

8 Temporal Mechanisms for Long-Term Memory

Lastly, a general theory of brain function needs to account for a second memory mechanism by which permanent, stable long-term memories are laid down and their contents retrieved. The nature of the storage mechanisms operant in brains, Lashley’s “engram” [80], is a fundamental problem that is intimately related to the neural coding problem. The competing theories of memory parallel those of neural coding and neural architectures [34, 101].

As Lorente observed, “permanent circulation of impulses in neural chains” cannot be the basis for long-term memory because these memory traces survive the cessation of neuronal activity that occurs under anesthesia, deep shock, and hibernation [102]. Like short-term memory, long-term memory is content-addressable. Neuronal activity associated with any attribute can be used to activate long-term memory traces that encode that attribute. Two alternative types of temporal long-term memory mechanisms can be envisioned that could store and retrieve events encoded via complex temporal spike patterns.

8.1 *Formation of Time-Delay Networks for Recognizing and Producing Temporal Patterns*

The first possibility tunes up time-delay networks to produce the complex temporal patterns that are to be stored. First the temporal patterns are repetitively presented to neuronal populations such that synapses with offsetting time delays that produce

spike-timing correlations within the local network are strengthened via spike-timing-dependent plasticity (STDP). The result would be that the repetitive complex patterns presented to the network would be assimilated in a manner similar to what was observed under the electrical conditioning experiments discussed earlier [58]. Subsequent activation of such a network might reproduce the consolidated delay pattern, thereby recreating the stored complex temporal pattern and injecting it into the rest of the network such that it can interact with signals that are currently circulating in regenerative short-term memory.

This explanation has the merit of relying on the same kinds of synaptic changes as connectionist theory. There is a great deal of accumulated biophysical evidence of semi-permanent changes in synaptic efficacy with use, and this process could involve those mechanisms. In effect, the neuronal assemblies become time-delay networks that are configured from past experience. These can become activated by incoming temporal patterns to facilitate channel activations (connectionist account) and/or to emit temporal patterns (timing net account).

8.2 *Polymer-Based Time-Space Molecular Memory for Storing Temporal Patterns*

A second possibility is that of a molecular memory that stores time patterns. Molecular memory mechanisms have been discussed for some time [103], and many are inspired by the power of the genetic nucleotide sequence code. A temporal molecular memory is attractive because it does not depend on particular, highly specific synaptic connections. Theories of RNA-based molecular memory (“memory RNA”) that were inspired by planaria memory transfer experiments were popular in the early 1960s, but due to failures to clearly replicate the basic transfer phenomena, this entire field was defunded at the NIH for a generation. In recent years, research on possible molecular memory mechanisms has been revived [64].

Complex temporal patterns lend themselves to instructional “tape recorder” memory mechanisms [33, 101, 104] that preserve temporal relations between events. If the attributes of the events are also temporally coded (e.g., sensory features of a particular place in a maze), then such temporal memory traces can serve as universal memory mechanisms whose form need not be radically transformed in the storage and retrieval process. I have previously proposed a mechanism similar to the scheme in Fig. 7 [105].

One potential molecular mechanism is that time patterns of intracellular ionic fluxes could be converted to spatial patterns of markers on polymers [106]. We

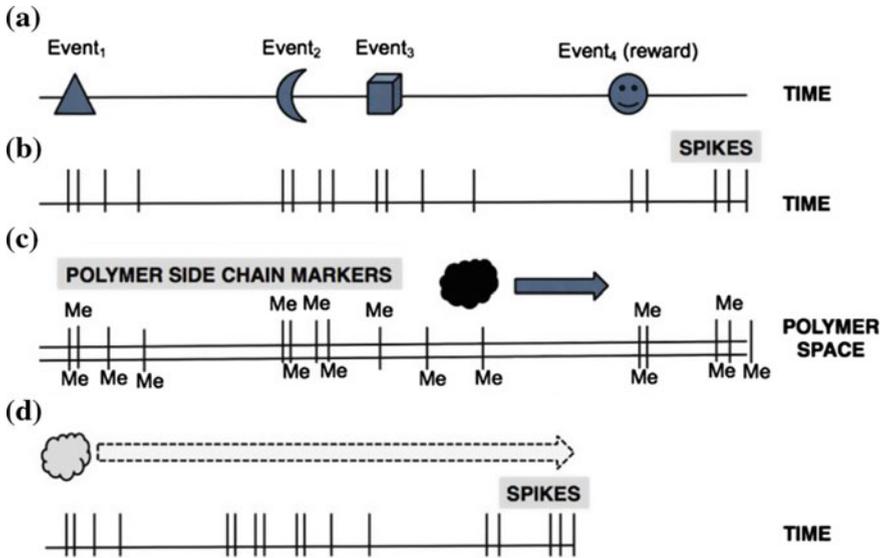


Fig. 7 Hypothetical scheme for long-term molecular storage of temporal patterns. **a** A sequence of temporally correlated events (1–3) that lead to a reward event (4). **b** Internal spike patterns produced by the internal events (e.g., encoding event-feature attributes). **c** Polymer (*double lines*) consisting of a backbone chain plus side-chains that can be chemically modified (e.g. methylated). A polymerase moves along the polymer backbone, adding a side-chain marker whenever the neuron is depolarized (or ionic concentrations change). The polymerase moves at a constant speed down the chain, thereby converting temporal pulse patterns to spatial patterns of markers on the polymer. **d** A second type of polymerase that moves down the chain at a faster rate, triggering ionic fluxes and generation of action potentials when a marked side chain is encountered. Such a mechanism would enable faster-than-real time readout that could subserve anticipatory steering of action

know that there are polymerases that move down the length of polymers, presumably at a constant average speed, and that there are also mechanisms for reversibly labeling the side chains of these polymers (e.g., methylation). Thus it is conceivable that temporal patterns of ionic fluxes (e.g., related to local calcium concentrations) could be laid down along the length of a polymer by a writing enzyme (there are much higher concentrations of DNA methylase in the nervous system than in other tissues). A second reading polymerase could scan the polymer at the same (or even faster) rate than the writing enzyme, thereby reading out the pattern. If all neural information related to relative event timings, attributes, and resulting reward or punishment is stored, as conditioning studies suggest, then a faster-than-real-time readout mechanism becomes a relatively simple means of predicting the future hedonic consequences (reward or punishment) of a present situation or course of action.

9 Conclusions

A theory of brain function based on complex temporal pattern spike codes, neural timing net architectures, and temporal memory mechanisms is outlined. Short-term temporal pattern memory entails regeneration of complex temporal patterns of spikes, whereas long-term temporal pattern memory could involve either tuning of time-delay networks or a polymer-based time-space molecular mechanism.

References

1. Rosen, R.: *Anticipatory Systems: Philosophical, Mathematical, and Methodological Foundations*. Pergamon Press, Oxford, New York (1985)
2. Rosen, R.: *Anticipatory Systems: Philosophical, Mathematical, and Methodological Foundations*, 2nd edn. Springer, New York (2012)
3. Nadin, M.: What speaks in favor of an inquiry into anticipatory processes? In: Klir, G. (ed.) *Anticipatory Systems: Philosophical, Mathematical, and Methodological Foundations*, pp. xv–lvii. Springer, New York (2012)
4. Louie, A.H.: Robert Rosen’s anticipatory systems. *Foresight* **12**, 18–29 (2010)
5. Nadin, M. (ed.): *Anticipation—Learning from the Past: The Russian/Soviet Contributions to the Science of Anticipation*, vol. 25. Springer, Cham, CH (2015)
6. Nadin, M. (ed.): *Anticipation Across Disciplines*, vol. 29. Springer, Cham CH (2016)
7. Tsagareli, M.G.: I.S. Beritashvili and psychoneural integration of behavior. In: Nadin, M. (ed.) *Anticipation: Learning from the Past: The Russian/Soviet Contributions to the Science of Anticipation*, vol. 25, pp. 395–414. Springer, New York (2015)
8. Vityaev, E.E.: Purposefulness as a principle of brain activity. In: Nadin, M. (ed.) *Anticipation: Learning from the Past: Russian/Soviet Contributions to the Science of Anticipation*, vol. 25, pp. 231–254. Springer, New York (2015)
9. Zueva, E.Y., Zuev, K.B.: The concept of dominance by A.A. Ukhomsky and Anticipation. In: Nadin, M. (ed.) *Anticipation: Learning from the Past: The Russian/Soviet Contributions to the Science of Anticipation*, vol. 25, pp. 13–35. Springer, New York (2015)
10. Rosenblueth, A., Wiener, N., Bigelow, J.: Behavior, purpose and teleology. *Philos. Sci.* **10**, 18–24 (1943)
11. Rosenblueth, A., Wiener, N.: Purposeful and non-purposeful behavior. *Philos. Sci.* **17**, 318–326 (1950)
12. de Latil, P.: *Thinking by Machine*. Houghton Mifflin, Boston (1956)
13. Ackoff, R.L., Emery, F.E.: *On Purposeful Systems*. Aldine-Atherton, Chicago (1972)
14. George, F.H., Johnson, L.: *Purposive Behaviour and Teleological Explanations*. Gordon and Breach Science Publishers, New York (1984)
15. Rashevsky, N.: *Mathematical Biophysics: Physico-Mathematical Foundations of Biology*, vols. I & II. Dover, New York (1960)
16. Rosen, R.: Biological systems as organizational paradigms. *Int. J. Gen. Syst.* **1**, 165–174 (1974)
17. Rosen, R.: *Life Itself*. Columbia University Press, New York (1991)
18. Kazansky, A.B.: Agental anticipation in the central nervous system. In: Nadin, M. (ed.) *Anticipation: Learning from the Past: The Russian/Soviet Contributions to the Science of Anticipation*, vol. 25, pp. 108–117. Springer, New York (2015)
19. Kilmer, W., McCulloch, W.S.: The reticular formation command and control system. In: Leibovic, K.N. (ed.) *Information Processing in the Nervous System*, pp. 297–307. Springer, New York (1969)

20. Mingers, J.: *Self-Producing Systems*. Plenum Press, New York (1995)
21. Maturana, H., Varela, F.: *Autopoiesis: the organization of the living*. In: Maturana, H., Varela, F. (eds.) *Autopoiesis and Cognition* (1980), vol. 42. D. Reidel, Dordrecht, Holland (1973)
22. Maturana, H.R.: *Autopoiesis*. In: Zeleny, M. (ed.) *Autopoiesis: A Theory of the Living*. North Holland, New York (1981)
23. Rosen, R.: Some realizations of (M, R) systems and their interpretation. *J. Math. Biophys.* **33**, 303–319 (1971)
24. Rosen, R.: What does it take to make an organism? In: Rosen, R. (ed.) *Essays on Life Itself*, pp. 254–269. Columbia University Press, New York (2000)
25. von Neumann, J.: The general and logical theory of automata. In: Jeffress, L.A. (ed.) *Cerebral Mechanisms of Behavior (the Hixon Symposium)*, pp. 1–41. Wiley, New York (1951)
26. Kauffman, S.: *The Origins of Order*. Oxford University Press, New York (1993)
27. Kampis, G.: *Self-Modifying Systems in Biology and Cognitive Science*. Pergamon Press, Oxford (1991)
28. Pattee, H.H., Raczaszek-Leonardi, J.: *Laws, language and life* Howard Pattee's classic papers on the physics of symbols with contemporary commentary by Howard Pattee and Joanna Raczaszek-Leonardi. *Biosemiotics* vol. 7. Springer, Dordrecht, New York (2012)
29. Cariani, P.: The semiotics of cybernetic percept-action systems. *Int. J. Signs Semiotic Syst.* **1**, 1–17 (2011)
30. Cariani, P.: Sign functions in natural and artificial systems. In: Trifonas, P.P. (ed.) *International Handbook of Semiotics*, pp. 917–950. Springer, Dordrecht (2015)
31. Graham, D.W.: *Aristotle's Two Systems*. Oxford University Press, New York (1987)
32. Modrak, D.K.: *Aristotle: The Power of Perception*. University of Chicago, Chicago (1987)
33. Favareau, D.: The evolutionary history of biosemiotics. In: Barbieri, M. (ed.) *Introduction to Biosemiotics*, pp. 1–67. Springer, Dordrecht (2008)
34. John, E.R.: *Mechanisms of Memory*. Wiley, New York (1967)
35. Milner, B., Squire, L.R., Kandel, E.R.: Cognitive neuroscience and the study of memory. *Neuron* **20**, 445–468 (1998)
36. Eichenbaum, H.: *The Cognitive Neuroscience of Memory: An Introduction*. Oxford University Press, New York (2012)
37. Eichenbaum, H.: Memory on time. *Trends in Cognitive Sciences* **17**, 81–88 (2013)
38. Atherton, L.A., Dupret, D., Mellor, J.R.: Memory trace replay: the shaping of memory consolidation by neuromodulation. *Trends Neurosci.* **38**, 560–570 (2015)
39. Snyder, B.: *Music and Memory*. MIT Press, Cambridge (2000)
40. Snyder, B.: Memory for music. In: Hallam, S., Cross, I., Thaut, M. (eds.) *The Oxford Handbook of Music Psychology*, pp. 107–117. Oxford University Press, Oxford, New York (2009)
41. Handel, S.: *Listening: An Introduction to the Perception of Auditory Events*. MIT Press, Cambridge, MA (1989)
42. Bregman, A.S.: *Auditory Scene Analysis, The Perceptual Organization of Sound*. MIT Press, Cambridge, MA (1990)
43. Fraisse, P.: *The Psychology of Time*. Harper & Row, New York (1978)
44. Longuet-Higgins, H.C.: *Mental Processes: Studies in Cognitive Science*. The MIT Press, Cambridge, MA (1987)
45. Longuet-Higgins, H.C.: A mechanism for the storage of temporal correlations. In: Durbin, R., Miall, C., Mitchison, G. (eds.) *The Computing Neuron*, pp. 99–104. Addison-Wesley, Wokingham, England (1989)
46. Thatcher, R.W., John, E.R.: *Functional Neuroscience, Vol. I. Foundations of Cognitive Processes*. Lawrence Erlbaum, Hillsdale, NJ (1977)
47. John, E.R., Bartlett, F., Shimokochi, M., Kleinman, D.: Neural readout from memory. *J. Neurophysiol.* **36**, 893–924 (1973)

48. Cariani, P.A., Delgutte, B.: Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase invariance, pitch circularity, rate pitch, and the dominance region for pitch. *J. Neurophysiol.* **76**, 1717–1734 (1996)
49. Cariani, P.: Temporal coding of periodicity pitch in the auditory system: an overview. *Neural Plasticity* **6**, 147–172 (1999)
50. Poeppel, D., Hickok, G.: Electromagnetic recording of the auditory system. *Handb. Clin. Neurol.* **129**, 245–255 (2015)
51. Will, U., Makeig, S.: EEG research methodology and brain entrainment. In: Berger, J., Turow, G. (eds.) *Music, Science, and the Rhythmic Brain: Cultural and Clinical Implications*, pp. xiv, 215 p. Routledge, New York (2011)
52. Arnal, L.H., Poeppel, D., Giraud, A.L.: Temporal coding in the auditory cortex. *Handb. Clin. Neurol.* **129**, 85–98 (2015)
53. Snyder, J.S., Large, E.W.: Gamma-band activity reflects the metric structure of rhythmic tone sequences. *Brain Res. Cogn. Brain Res.* **24**, 117–126 (2005)
54. Trainor, L.J., Zatorre, R.: The neurobiological basis of musical expectations. In: Hallam, S., Cross, I., Thaut, M. (eds.) *The Oxford Handbook of Music Psychology*, pp. 171–183. Oxford University Press, Oxford, New York (2009)
55. Koelsch, S.: *Brain and Music*. Wiley-Blackwell, Chichester (2012)
56. Malmierca, M.S., Sanchez-Vives, M.V., Escera, C., Bendixen, A.: Neuronal adaptation, novelty detection and regularity encoding in audition. *Front Syst. Neurosci.* **8**, 111 (2014)
57. John, E.R.: Electrophysiological studies of conditioning. In: Quarten, G.C., Melnechuk, T., Schmitt, F.O. (eds.) *The Neurosciences: A Study Program*, pp. 690–704. Rockefeller University Press, New York (1967)
58. Morrell, F.: Electrical signs of sensory coding. In: Quarten, G.C., Melnechuk, T., Schmitt, F.O. (eds.) *The Neurosciences: A Study Program*, pp. 452–469. Rockefeller University Press, New York (1967)
59. Schultz, W., Dickinson, A.: Neuronal coding of prediction errors. *Annu. Rev. Neurosci.* **23**, 473–500 (2000)
60. Schultz, S.R., Panzeri, S.: Temporal correlations and neural spike train entropy. *Phys. Rev. Lett.* **86**, 5823–5826 (2001)
61. Miller, R.R., Barnet, R.C.: The role of time in elementary associations. *Curr. Dir. Psychol. Sci.* **2**, 106–111 (1993)
62. Savastano, H.I., Miller, R.R.: Time as content in Pavlovian conditioning. *Behav. Processes.* **44**, 147–192 (1998)
63. Arcediano, F., Miller, R.R.: Some constraints for models of timing: a temporal coding hypothesis perspective. *Learn. Motiv.* **33**, 105–123 (2002)
64. Tucci, V., Buhusi, C.V., Gallistel, R., Meck, W.H.: Towards an integrated understanding of the biology of timing. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **369**, 20120470 (2014)
65. Boring, E.G.: *Sensation and Perception in the History of Experimental Psychology*. Appleton-Century-Crofts, New York (1942)
66. Uttal, W.R. (ed.): *Sensory Coding: Selected Readings*. Little-Brown, Boston (1972)
67. Uttal, W.R.: *The Psychobiology of Sensory Coding*. Harper and Row, New York (1973)
68. Rieke, F., Warland, D., de Ruyter, R., Steveninck, V., Bialek, W.: *Spikes: Exploring the Neural Code*. MIT Press, Cambridge, MA (1997)
69. Cariani, P.: As if time really mattered: temporal strategies for neural coding of sensory information. Reprinted in: K Pribram, ed. *Origins: Brain and Self-Organization*, Hillsdale, NJ: Lawrence Erlbaum, 1994; 208–252. 12, 161–229 (Reprinted in: K Pribram, ed. *Origins: Brain and Self-Organization*, Hillsdale, NJ: Lawrence Erlbaum, 1994; 1208–1252) (1995)
70. MacKay, D.M.: Self-organization in the time domain. In: Yovitts, M.C., Jacobi, G.T., Goldstein, G.D. (eds.) *Self-Organizing Systems 1962*, pp. 37–48. Spartan Books, Washington, D.C. (1962)
71. Pratt, G.: Pulse computation. Department of Electrical Engineering and Computer Science, vol. Ph.D., pp. 214 leaves. Massachusetts Institute of Technology, Cambridge, MA (1989)

72. Braitenberg, V.: The neuroanatomy of time. In: Miller, R. (ed.) *Time and the Brain*, pp. 391–396. Harwood Academic Publishers, Australia (2000)
73. Abeles, M.: Synfire chains. In: Arbib, M.A. (ed.) *The Handbook of Brain Theory and Neural Networks* (2nd Ed.), pp. 1143–1146. MIT Press, Cambridge, MA (2003)
74. Izhikevich, E.M.: Polychronization: computation with spikes. *Neural Comput.* **18**, 245–282 (2006)
75. Cariani, P.: Temporal coding of periodicity pitch in the auditory system: an overview. *Neural Plast.* **6**, 147–172 (1999)
76. Cariani, P.: Temporal coding of sensory information in the brain. *Acoust. Sci. Tech.* **22**, 77–84 (2001)
77. Perkell, D.H., Bullock, T. H.: Neural coding. *Neurosciences Research Program Bulletin.* **6**, 3 221–348 (1968)
78. Emmers, R.: *Pain: A Spike-Interval Coded Message in the Brain*. Raven Press, New York (1981)
79. Cariani, P.A., Delgutte, B.: 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. Neurophysiol.* **76**, 1698–1734 (1996)
80. Orbach, J.: *The Neuropsychological Theories of Lashley and Hebb*. University Press of America, Lanham (1998)
81. Cariani, P., Micheyl, C.: Towards a theory of information processing in the auditory cortex. In: Poeppel, D., Overath, T., Popper, A. (eds.) *Human Auditory Cortex: Springer Handbook of Auditory Research*. Springer, New York (2012)
82. Zanto, T.P., Snyder, J.S., Large, E.W.: Neural correlates of rhythmic expectancy. *Adv. Cogn. Psychol.* **2**, 221–231 (2006)
83. Fujioka, T., Trainor, L.J., Large, E.W., Ross, B.: Internalized timing of isochronous sounds is represented in neuromagnetic beta oscillations. *J. Neurosci.* **32**, 1791–1802 (2012)
84. Nozaradan, S.: Exploring how musical rhythm entrains brain activity with electroencephalogram frequency-tagging. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **369** (2014)
85. Hickok, G., Poeppel, D.: Neural basis of speech perception. *Handb. Clin. Neurol.* **129**, 149–160 (2015)
86. Ding, N., Melloni, L., Zhang, H., Tian, X., Poeppel, D.: Cortical tracking of hierarchical linguistic structures in connected speech. *Nat. Neurosci.* (2015)
87. Thaut, M.: *Rhythm, music, and the brain: scientific foundations and clinical applications*. Routledge, New York (2005)
88. Large, E.W., Snyder, J.S.: Pulse and meter as neural resonance. *Ann. N. Y. Acad. Sci.* **1169**, 46–57 (2009)
89. Doelling, K.B., Poeppel, D.: Cortical entrainment to music and its modulation by expertise. *Proc. Natl. Acad. Sci. USA* **112**, E6233–E6242 (2015)
90. Cariani, P., Micheyl, C.: Towards a theory of information processing in the auditory cortex. In: Poeppel, D., Overath, T., Popper, A. (eds.) *Human Auditory Cortex: Springer Handbook of Auditory Research*, pp. 351–390. Springer, New York (2012)
91. Doelling, K.B., Arnal, L.H., Ghitz, O., Poeppel, D.: Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. *Neuroimage* **85**(Pt 2), 761–768 (2014)
92. Farbood, M.M., Rowland, J., Marcus, G., Ghitz, O., Poeppel, D.: Decoding time for the identification of musical key. *Atten. Percept. Psychophys.* **77**, 28–35 (2015)
93. Swanson, L.W.: *Brain Architecture: Understanding the Basic Plan*. Oxford University Press, New York (2012)
94. Cariani, P.: Neural timing nets. *Neural Netw.* **14**, 737–753 (2001)
95. Cariani, P.: Temporal codes, timing nets, and music perception. *J. New Music Res.* **30**, 107–136 (2002)
96. Cariani, P.A.: Temporal codes and computations for sensory representation and scene analysis. *IEEE Trans Neural Netw./Publication IEEE Neural Netw. Council.* **15**, 1100–1111 (2004)

97. Cariani, P.: Outline of a cybernetic theory of brain function based on neural timing nets. *Kybernetes* **44**, 1219–1232 (2015)
98. Sporns, O.: *Networks of the brain*. MIT Press, Cambridge, Mass (2011)
99. Abeles, M.: *Local Cortical Circuits. An Electrophysiological Study*. Springer, Berlin (1982)
100. Abeles, M.: *Corticonics*. Cambridge University Press, Cambridge (1990)
101. John, E.R.: Switchboard vs. statistical theories of learning and memory. *Science* **177**, 850–864 (1972)
102. Lorente de No, R.: Circulation of impulses and memory. In: Schmitt, F.O. (ed.) *Macromolecular Specificity and Biological Memory*, pp. 89–90. MIT Press, Cambridge, MA (1962)
103. Schmitt, F.O.: Biologically structured microfields and stochastic memory models. In: Schmitt, F.O. (ed.) *Macromolecular Specificity and Biological Memory*, pp. 7–17. MIT Press, Cambridge, MA (1962)
104. John, E.R.: Studies of memory. In: Schmitt, F.O. (ed.) *Macromolecular Specificity and Biological Memory*, pp. 80–85. MIT Press, Cambridge, MA (1962)
105. Cariani, P.: Symbols and dynamics in the brain. *Biosystems* **60**, 59–83 (2001)
106. Landry, C.D., Kandel, E.R., Rajasethupathy, P.: New mechanisms in memory storage: piRNAs and epigenetics. *Trends Neurosci.* **36**, 535–542 (2013)

Time is of the essence

Proceedings of A Body of Knowledge - Embodied Cognition and the Arts conference CTSA UCI 8-10 Dec
2016

Peter Cariani

I. Introduction

This paper outlines two ideas. The first proposes a basic high-level neuropsychological and neurophenomenological cybernetic framework for discussing the structure of mind and experience. The second is that much, perhaps even most, informational processes in the brain are inherently temporal in nature, i.e. that they are subserved by temporal neural codes. To paraphrase Mari Reiss Jones, in the study of mind and brain, “time is our lost dimension” (Jones 1976). In this view, there is pervasive, common temporal structure in the internal neural representations that subserve both perception and action. This common temporal structure permits perception to facilitate, inform, and even bootstrap action, and vice versa. Time structure in perception, action (movement, behavior), cognition, affect, motivation (drives, goals), and memory may allow these different mental faculties to mutually influence one another.

II. A functionalist psychological framework

A basic neuropsychological framework can be laid out that consists of complementary semiotic (informational) and systems functionalities: sensing (of external and internal states), effecting, coordination (mapping of sensory states to motor-action states), memory (short and long term), cognition (pattern analysis, recognition, and generation), affect, motivation (needs, drives, goals) and switching of internal modal behavioral states. The framework comports with much of standard psychology. The framework is *cybernetic* in that it conceives of minds and brains in terms of purposive, purposeful systems (Craik 1966; Ackoff and Emery 1972; DeYoung 2015; Powers 1973) whose behavior is mostly driven by embedded goal-seeking feedback mechanisms (Rosenblueth and Wiener 1950; Rosenblueth et al. 1943; de Latil 1956).

All organisms are engaged in self-maintenance (homeostasis) and continual self-construction that realizes a self-production organizational network, an *autopoiesis* (Maturana and Varela 1980). All of them are embedded in external environments and are constantly reacting and responding to sensed changes in it. Animals are organizationally closed by virtue of their self-production and construction, but informationally open by virtue of inputs that they do not control.

Semiotic systems. An *informational system*, in the sense used here, is a system whose behavior is mediated by sign-distinctions, i.e. pattern- or configuration-based switchings, such that its behavior can be effectively described and predicted in such terms. This contrasts with systems and processes, such as digestion, that do not enable such semiotic descriptions. Whereas the behavior of an informational system can be effectively described both in terms of sign-distinctions and physical processes (e.g. dynamical systems models), some systems (rivers) and processes (digestion) are only effectively described in terms of dynamical systems.

Sign-mediated percept-action systems. Animals with nervous systems are semiotic systems because their behavior can be effectively described in terms of sign-distinctions in the form of neural spike train signals. Neural signals are patterns of activity that are involved in informational functions. To paraphrase Bateson these are different patterns of spikes that make a difference in terms of functional states and behaviors. Types of patterns that make specific functional differences are called neural codes (in the same way that nucleotide sequences that are functionally meaningful to transcription-translation systems constitute the genetic code).

All organisms, including those without nervous systems, to some degree act on the basis of distinctions (signs) in that they make discrete behavioral choices. The ability to switch between alternative behaviors dramatically increases in complexity and effectiveness with the evolution of nervous systems in animals. Animals are motile organisms that do not produce their own source

of food (stored energy). As such, animals must depend, directly or indirectly, on organisms (algae, plants) that are capable of collecting and converting energy into storable form. Being motile, they must navigate their environments and encounter different local situations. This in turn necessitates sensing the environment — making distinctions — and acting in a flexible manner that is contingent on current distinguished situations and conditions.

Nervous systems are specialized internal signaling systems that provide 1) sensory functions that transduce environmental states and events into neural signals, 2) coordinative functions (communication, command, control) that regulate internal bodily functions and organize behavior, and 3) effector functions (muscles, secretory organs) that operate on the realm outside the body (action). These functions are schematized in Figure 1.

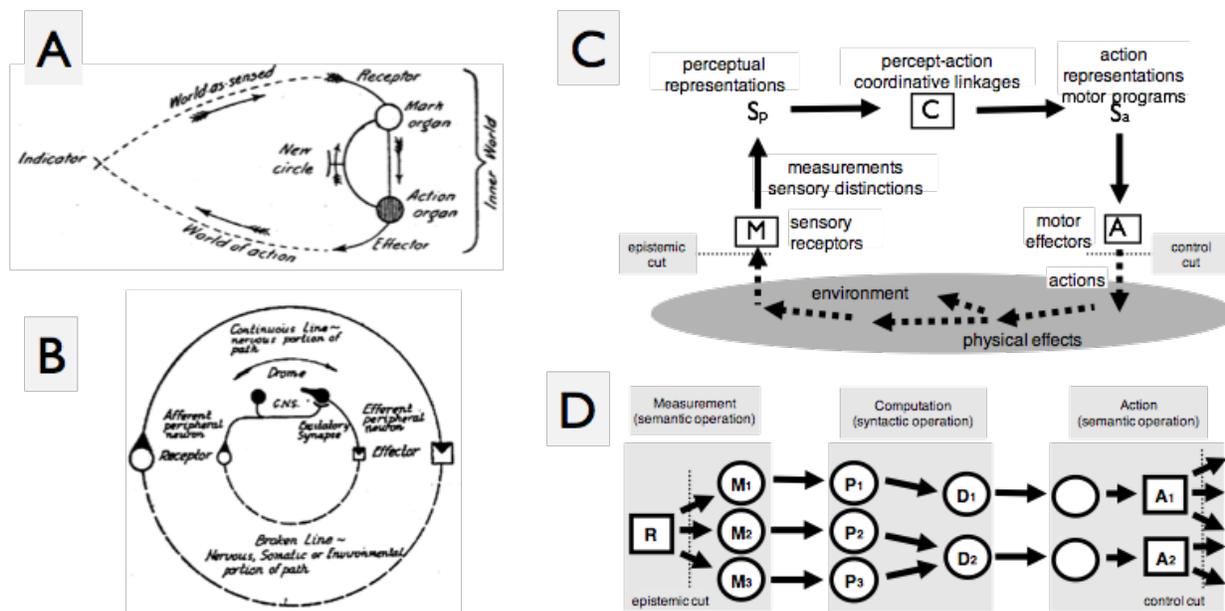


Figure 1. Percept-coordination action cycles and state-transition structure. A. Uexküll's percept-action and internal cycles that define an organism's *umwelt* (Uexküll 1926). B. McCulloch's circular causal percept-action paths that run within the nervous system and through the environment (McCulloch 1946). C. Operational structure of percept-action cycles (Cariani, 1989, 2001, 2011, 2015). D. State transition structure of a percept-coordination-action sequence. Measurements, coordinations ("computations"), and actions have fundamentally different operational state-transition structures. Here "computation" means reliable, determinate sequences of functional states (each predecessor state has one successor state). Sensing involves registering the outcomes of contingent interactions with the external environment (one predecessor reference state R to one of several possible successor states M₁, M₂, M₃), whereas action involves producing changes in the environment. Epistemic and control cuts define functional boundaries between contingent and determinate state transitions that distinguish an internal realm of reliable (self) control from an external realm of uncontrolled processes (the environment).

Sensory functions ("measurements") involve making distinctions that reduce uncertainty vis-a-vis (perceived) environmental states (Ashby's concept of "information" (Ashby 1956)). Effector systems (muscles, secretory organs, electric organs) cause environmental changes ("actions"). Coordinative functions involve choosing which actions are to be carried out given different perceived environmental contexts. Coordinative functions in higher animals include processes of perception, cognition, conation, memory, imagination/simulation, deliberation, affect, and control of behavioral mode and systems state (e.g. waking-sleeping-hypnotic state).

It has been argued at length elsewhere (Cariani 1989, 2001b, 2011; Cariani 2015b) that the basic functionalities of sensing, coordinating, and effecting can be clearly, operationally distinguished

on the basis of the state-transition structure of functional states (contingent vs. determinate, externally- vs. internally-driven). This demarcation (Fig. 1 bottom, dotted lines) depends on choosing an appropriate observational frame, i.e. the set of observables through which the behavior of the system is viewed.¹

The clearest examples lie in the operational structure of scientific models, where we can distinguish between ill-defined, contingent measurement operations and well-defined formal symbolically mediated processes of calculation and prediction. If one adds effectors and uses the predictive apparatus to drive them, one has a robotic devices with discrete internal states that can act on its environment. Although systems with identified discrete functional states are easiest to describe in these operational terms, the methodology can also be applied to analog systems that have attractor basins in their dynamics (alternate modes).²

Percept-action cycles. In waking, acting states, animals are constantly engaging in transactions with their environments. Effective action entails ongoing, iterated cycles of perception, coordination, and action. Because these are circular-causal processes, one can view action as being contingent on perception and coordination, or perception being contingent on action (Powers 1973).

Sensory systems. Animals have an array of means of sensing states and events inside their bodies and in the world outside their bodies. Internal senses, such as body position (proprioception, kinesthesia, posture), muscle tension, balance, pain, hunger/satiety, thirst, and emotions provide a read-out on the state of the organism. External senses, such as vision, audition, touch, temperature-sense, smell, and taste are directed towards realms outside the body. Sensory systems interact with the external environment and subsequently register one of several possible alternative sensory states, contingent on the current state of the environment. The process of reading out one definite outcome from many possible ones provides “information” related to the sensor-environment interaction that reduces uncertainty. Because of its operational state-transition structure and the functions that it affords, sensing is fundamentally different from coordination or action, in the same way that measurements play a fundamentally different role in predictive scientific models from formal computations.³

¹ There is no “correct” or “incorrect” frame, only those that are pragmatically more or less useful for a given predictive, explanatory, or practical purpose. Those interested in dissolving boundaries and functional distinctions will choose one set of observables (e.g. low-level physical dynamics), whereas those interested in clearly drawing boundaries and distinctions can choose those frames (e.g. neural signals, neural representations) that suit those purposes. For most animals with nervous systems these distinctions can be clearly drawn because the requisite underlying structure exists, but for most inanimate material systems (e.g. the sun, clouds, rivers), it may be impossible to find any frame that will allow the system to be seen as autopoietic, semiotic (sign-driven), purposeful, or anticipatory because the internal relations that support these functional organizations simply don’t exist in those systems.

²The attractor basins of the dynamicists are the symbols of the computationalists. There is a third path that incorporates the strengths of both paradigms. I have argued for intermediate level analog signal descriptions of brains that involves an internal dynamics of mixed analog-iconic and digital-symbolic neural signals (Cariani 2001b; Cariani 2015a, 2017), cf. intermediate level neural signal dynamics of (Prinz 2012). The question is not whether mind/brain functions are to be described either in terms of symbols, signals, or dynamics, but which of these types of descriptions best fits for what one wants to explain. This is in keeping with Pattee’s notion of complementarity between different modes of description (Pattee and Raczaszek-Leonardi 2012).

³ Processes of measurement (sensing) are external to formal systems. In terms of Turing’s theory of deterministic automata, any contingent inputs from without are not part of the automaton, but constitute arbitrary inputs from “oracles” whose origins are ill-defined from the perspective of the formal description. Similarly, measurements and formal computations play categorically distinct roles in scientific predictive models, as von Neumann argued in the 1930’s debate over the foundations of physics.

Motor systems. Effector systems realize actions. Effectors act on the external world to change it in some way. Almost all actions involve voluntary movements that are mediated by muscles. These voluntary movements permit animals to move around their environments, obtain food and water, eat, drink, defend themselves, and mate. Effecting also includes non-motoric actions, such as secretions (e.g. ink, venom, digestive juices, pheromones, sperm) or production of electric fields for stunning prey (electric eel).

Coordinative systems. Coordinative systems choose which actions to pursue given current perceived affordances (conditions and opportunities), goal-related drive-states, inherited predispositions, and estimations of effectiveness based on previous experience. The coordinative functionality of Fig. 1 would include the goal selection and steering processes of Fig. 2 and all of the circular-causal loops in Fig. 3.

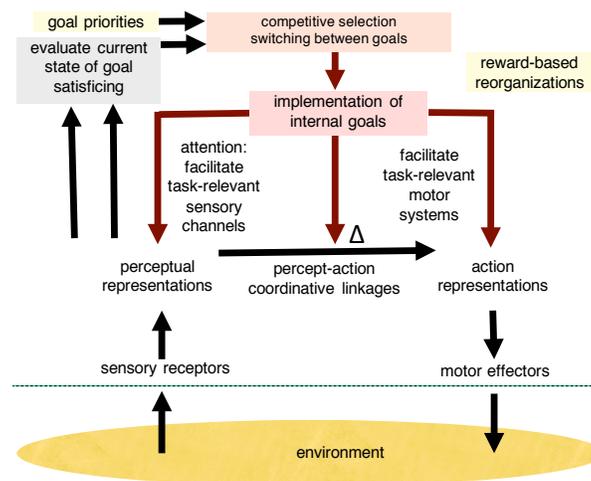


Figure 2. Embedded goals and percept-action cycles. Evaluative processes and competitive goal selection mechanisms switch between dominant goals. Sensory channels, percept-action mappings, and motor routines relevant to current goals are facilitated, whereas others are inhibited. Reward mechanisms reorganize task-relevant switchings so as to increase the probabilities of actions that achieve rewards and/or avoid punishments.

Purposive systems. Animals are purposive, purposeful systems in that they are feedback-control systems whose behavior is guided by embedded goal-states. The goals range from low-level system-imperatives for homeostasis (maintain adequate oxygenation, hydration, ionic balances, blood-sugar, adenosine clearance via sleep) to higher level imperatives important for survival (find food and water, avoid/flee/fight predators) and reproduction (courtship, mating) and beyond. Maslow's hierarchy of needs further extends these goals to safety, security, social status and self-actualization.

The goals are in constant competition for control over the behavior of the animal, but they are defeasible constraints such that their priorities are not fixed, but depend on several factors. These include drive state (how well a given goal is being realized or satisfied and how urgent is the goal, i.e. how important is it to pursue *this goal now*), perceived opportunity for achieving the goal (what goal-related affordances are perceived to exist in the current situation) and what are the perceived odds of their success (based on past experience).

Purposive systems thus have internal evaluative measures of how well goals are being met as well as mechanisms that mediate the competition between competing goals and that bias the behavior of the system towards the realization of those goal-choices that are dominant at any given moment.

In Fig. 3 this competitive process is depicted goal-dynamics. The Russian psychologist Ukhtomsky (1845-1942) proposed such a “dominance principle” for behavior (Zueva and Zuev 2015):

“... in the nervous system, there is at each moment only one active dominating dynamic structure or constellation of excitation, which is associated with the most actual, urgent ongoing needs and desires. This excitation structure plays the role of a situational nervous network, an agency for organizing the physiological and behavioral response directed to satisfying these needs. At the same time all other goals and desires are suppressed.” (Kazansky 2015)

“What is the difference between a bodily physiological mechanism in an animal and a technical mechanism? Firstly, the former is generated during the course of the reaction itself. Secondly, once chosen, the behavior of a technical mechanism is secured once and for all by its construction, whereas in a reflex apparatus it’s possible to successively realize as many different mechanisms as there are available degrees of freedom in the system. Each of the successively realized mechanisms is achieved through the active inhibition of all available processes, except one.” A.A. Ukhtomsky, quoted in(Kazansky 2015)

Goal-dynamics in brains. The notion of competitive goal dynamics fits current conceptions of brains as sets of competing circuits that steer behavior in service of different purposes (goals, end-states, drives, desires, internal imperatives). The inhibitory suppression all goal-related behavioral patterns save the one being currently realized fits with emerging theories of the basal ganglia/striatum as enabling and facilitating those motor and sensory circuits needed to perform specific tasks. Below each patch of cerebral cortex is an associated stratum of basal ganglia. This circuitry involves an inhibitory-inhibitory pathway that acts as a generalized double-inhibition braking system (increasing the activity of the first inhibitory link decreases the activity of the second inhibitory link, thereby disinhibiting (facilitating) the target neural populations. Brakes thus can be selectively released to facilitate thalamic sensory and motor neural subpopulations that are relevant to a particular task (the presently-dominant goal). On the motor side, this facilitates particular neural assemblies that subserve task-specific actions, whereas on the sensory side, it facilitates particular sensory channels carrying task-relevant information, thereby realizing a neural gain-control mechanism for attention.

In the normal resting state, inhibition makes neural loop-gains slightly attenuating, such that incoming signal patterns rapidly die out. When a brake is applied, the ensuing disinhibition changes these loop-gains to slightly amplifying, such that neuronal sensory and motor signals related to the current goal become the predominant signals within the system. It is as Ukhtomsky said: "Each of the successively realized mechanisms is achieved through the active inhibition of all available processes, except one" (Kazansky, 2015).

Adaptive systems. Systems are adaptive if their structures and functions can be altered, on the basis of past and present experience, in such a way that performance vis-a-vis embedded goals is improved (goals are better or more reliably attained). In percept-coordination-action systems, percept-action mappings can be altered contingent on experience. There are two different kinds of adaptive processes, two modes of creating novelty (Cariani 2012). In *combinatorial adaptivity* new mappings between existing percepts and actions are chosen. In *creative adaptivity* emergent novelty is generated by producing new primitives. For example, entirely new types of perceptual distinctions can be created by evolving structures that enable new sensory distinctions, such as the ability to distinguish additional colors.

Functional and phenomenal boundaries. Sensors and effectors create functional boundaries between the inner states of the percept-action system that in Üexkull’s terms constitute its *umwelt* or life-world and the external environment (Fig. 1A). The boundaries reflect the structure of

contingent vs. determinate and externally-driven vs. internally-driven circular-causal changes. The epistemic and control cuts demarcate an internal realm of self-produced and self-controlled activity (organizational closure) vs. an external realm that is outside the full control of the percept-action system. Percept-action systems are thus *informationally-open* (to environmental changes that are registered by sensors), but *organizationally-closed* (they determine their own internal structure and choose which sensors on which they base their current behavior). When a percept-action system gains the ability to fully control some part of the world outside of its sensors, effectors, and coordinative parts, that newly controlled part of the world becomes part of the realm of the functional boundaries of that system move outward.

Phenomenologically, the boundaries created by these contingencies appear to correspond to our sense of self (what is under our control, part of ourselves) and non-self (what is independent of our control, separate from ourselves). Addition of a prosthesis pushes the realm of control outward from our physical bodies such that the epistemic or control cut moves to the point of contingency, which now lies at the distal end of the prosthetic device.

Anything that can be attached to and reliably controlled by a percept-action system, such as devices that augment our senses (microscopes, telescopes, geiger counters), our cognition (computer programs, the internet), our memory (pencil-and-paper memos, the memories of other people) or our motor systems (tools, cars and other vehicles) can function as a prosthetic device that enables extended perception, cognition, memory, and action. The construction of external prosthetic devices extends the mind beyond the boundaries of the biological body.

Functional and phenomenal boundaries can also move inwards into the nervous system by means of iconic neuronal representations that preserve the correlation structure of the sensory flux (Cariani 2012). By virtue of the temporal correlations between stimulus structure and spikes, iconic temporal codes for pitch, rhythm, visual form, and optic flow bring the form of sensory flux into the nervous system. Neural assemblies that recognize particular forms then function as internal sensory surfaces that can subsequently switch behavior. It is thus possible to think of the formation of new neural assemblies in terms of creating new concept-primitives, which serve as new windows on the internal milieu of iconic neuronal representations that in turn carries the correlation structure of the external world within it.

Organization embedded in matter. All of these functionalities are due to relational organizations embedded in material systems. The working ontology is a form of Aristotelian hylomorphism, which sees functional properties as dependent on organization that is embedded in matter (Graham 1987; Modrak 1987).

To summarize these organizational dependencies, animals are *living systems* by virtue of the active regeneration of their organization (material components and relations), conceived in terms of self-production, autopoiesis, metabolism-repair, kinematic self-reproduction, autocatalytic nets), self-modifying systems, and/or semantically-closed self-interpreting construction systems. They are *semiotic systems* because their internal operation can be effectively described in terms of sign-distinctions conveyed via neural pulse codes. They are *informational systems* in that their behavior is mainly determined by distinctions rather than through energetic mechanisms. They are *autonomous purposive systems* because they are mainly driven by internal goals. Animals have their own *agency* to the extent that they have embedded goals (internal motivation), requisite ability (the right mechanisms needed for action) and sufficient freedom of action (autonomy) to reliably achieve particular goals. Animals are *anticipatory systems* if they have learning and memory mechanisms that allow them to project the past into the present so as to predict future consequences of current courses of action. They are *adaptive* because they can modify themselves

in a manner that improves their performance vis-a-vis evaluative criteria (survival, reproduction, health). They are *emergent* both because they can create both new combinations of existing structural and functional primitives and because in some cases they can create new primitives as well.

If one defines these different attributes in terms of these specific, distinguishable kinds of *functional organizations*, then a system, such as an autonomous robot, need not be living to be semiotic, autonomous, purposive/agentive, anticipatory, and/or adaptive. In this view, because these functional attributes depend on different aspects of organizations, although biological organisms have all of these attributes, it is not necessarily the case that *living = purposive = semiotic = anticipatory = adaptive = emergent* (see discussion of teleology and autonomy in (Thompson 2007)).

III. A neuropsychological and neurophenomenal framework

Neuropsychology involves the relation of neural processes to mental operations and functions. A more explicitly neuropsychological framework for mental operations and neural processes is shown in Figure 3. It includes operations related to perception (sensations), cognition (recognitions, thoughts), emotion (feelings), conation (goals, drives), memory (short- and long-term), global state control (wake-sleep cycles), and mechanisms for switching between behavioral programs associated with different goal states (action-switching and attention). Although they constantly interact and influence one another, the processes of perception, cognition, emotion, conation, memory, and the execution of action nevertheless involve distinct and relatively autonomous *types* of operations.

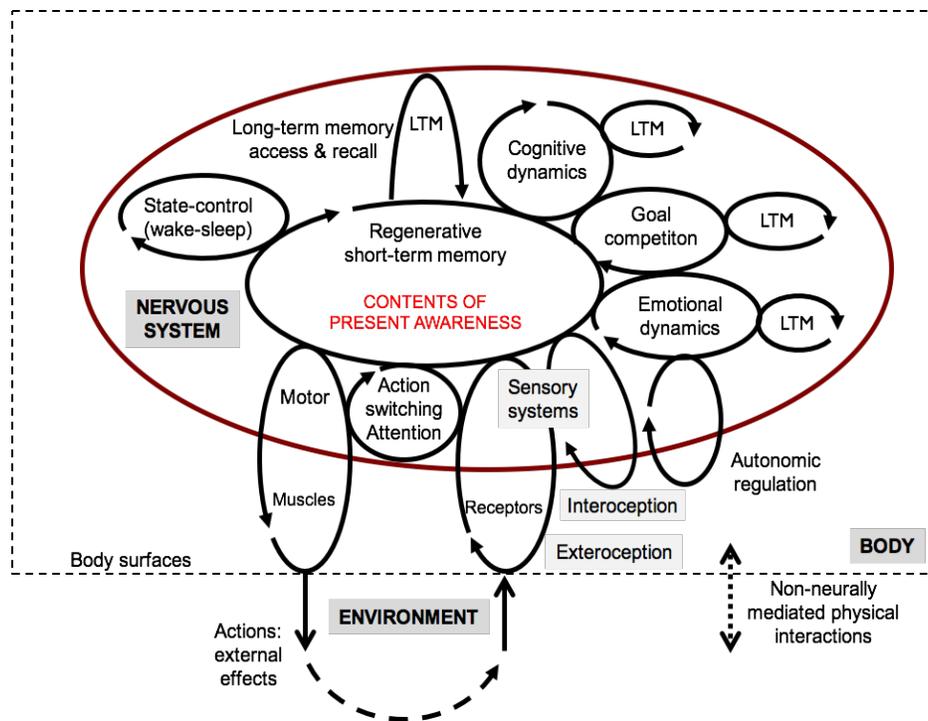


Figure 3. Mind-brain as a network of circular-causal processes. The three realms of environment, body, and nervous system are outlined. Within each of the loops, alternative sets of competing neural signals related to different types of distinctions (sensory, motor, cognitive, affective, conative, state-associated) build-up, compete, and are actively regenerated or die out. Those signals that are actively regenerated in global short-term memory are hypothesized to

form the contents of conscious awareness at any given moment. The model is signal-centric (vs. connection-centric) in that the signals are assumed to be complex, higher-dimensional temporal patterns of spikes that interact on selected dimensions, such that signal dynamics realizes most informational functions.

The schematic demarcates boundaries between environment, body, and nervous system. Each functionality itself has many subdivisions that are not shown. For example, the sensory functionality would include exteroceptive senses such as vision, audition, touch, temperature, smell, taste and interoceptive senses such as pain. In addition to the operations discussed above in the context of percept-coordination-action cycles, this framework explicitly adds short-term reverberating, dynamically regenerative memory and long-term memory storage and retrieval processes, as well as neural circular-causal loops that mediate goal-competition (conation), affect, cognitive processing, regulation of system state (wake-sleep-hypnosis), and the switching of attention. This more detailed framework comes closer to a high-level model of animal minds and mind-body-environment relations than the more abstract and general framework of percept-coordination-action.

Memory and anticipation. In addition to embedded goals, percept-action systems can also have memory mechanisms, which enable vastly more efficient adaptive learning mechanisms. Systems without memory are reactive, they are determined by their immediate inputs, whereas the behavior of systems with memory depends not only on current outputs but also on their past history. Memory involves the making of records that persist in time such that they can be read-out and utilized at future times. All organisms have a genetic memory embedded in their DNA that contains sequences of nucleotides that specify the construction of specific proteins.

In percept-action systems, memory permits action to be made contingent on temporal sequences and patterns of perceived events rather than on only current perceptual states. Memory also permits actions to be sequenced and timed to form complex patterns of linked actions. The ability to handle temporal patterns of sensations and actions dramatically increases the dimensionality of percept-action spaces. Memory enables anticipation of future events and rewards on the basis of past experience combined with current conditions. Memory mechanisms that record sequences of internal state transitions can be used as templates for guiding behaviors that increase likelihoods for reward and decrease those for punishment (Cariani 2017).

System modes. Almost all animals have cycles of activity and inactivity associated with wake-sleep cycles. Many animals exhibit short-lived quasi-hypnotic states (birds, reptiles) and orgasmic states. Humans also have various system-states that are associated with different states of consciousness: e.g. hypnagogia, hypnosis, trance, meditation, ecstatic spiritual states.

Behavioral modes. Animals engage in a relatively limited number of behavioral modes that include basic activities such as drinking, eating, eliminating, fleeing, fighting, grooming, courting, mating, resting, exploring, playing (Kilmer and McCulloch 1969). Many, but not all, of these modes, involve actions.

Neural implementation. Each functionality is subserved by sets of recurrent neuronal circuits. The brain is a network of recurrent, cyclic paths, i.e. circuits, “loops” that realize circular-causal feedback processes. For example, the neural interconnections between thalamus and cortex and between cortical regions are reciprocal — for every “forward” projection there is a re-entrant “reverse” projection. The hippocampus itself contains an array of recurrent paths that have been thought to implement an autoassociative memory.

In the brain different types of structures appear to perform different classes of functions. Roughly, the cerebral cortex performs pattern-analysis, recognition, and generation; the striatum switches between motor programs and attentional facilitations of sensory information according to

immediate goals (tasks); midline dopamine circuits engage in reward evaluation, predictive timing, and adaptive reorganization of neural circuits; the hippocampus rebroadcasts neural signals related to hedonically salient events to enable formation of long term memories; the cerebellum carries out fine timing adjustments; the amygdala carries out threat assessments of novel stimuli.

Different regions of the cerebral cortex realize different functions mainly due to the correlational structure of their inputs, which depends on how they are situated in relation to sensory and motor surfaces and to other cortical regions. However, a given patch of cortex may realize multiple functions simultaneously (e.g. encoding multiple attributes of a sound or visual event), and regions of cortex deprived of their normal inputs (such as the visual cortex in a blind person) can eventually be repurposed to subserve other kinds of perceptual and cognitive functions. It now appears that music perception and cognition activate most of the same cortical territories as speech and language functions. All of this casts doubt on the kinds of extreme modularities and dedicated functional specificities that some neurocognitivists ascribed to particular cortical regions.

Neurophenomenology. Neurophenomenology involves the relation of neural processes to the structure of experience, both the requisite conditions for having an experience (NCCs, “neural correlates of consciousness”, (Koch 2004), i.e. neural requisites for states of consciousness) and the specific conditions that produce a specific experience (NCCCs, “neural correlates of the contents of consciousness”, i.e. the neural concomitants of qualia). Neurophenomenological theories have always existed alongside neurological, neurophysiological and neuropsychological investigations (Boring 1933, 1942; Troland 1929; Fessard 1954; Köhler 1947) as theorists have attempted to relate the structure of neural activity with that of their own subjective experience, e.g. (Varela 1996).

This framework assumes that neural signals, however structured, circulate in the various functional circuits that subserve the different functionalities. My working hypothesis is that the phenomenal concomitants of sets of neurally-coded signals that are actively regenerated in global loops at any given moment constitute the contents of awareness. Further, the contents of awareness at any given moment is isomorphic to the specific neural signals that are present in those loops.

Neural codes. The nature of the neural signals circulating in the circuits is critical. The neural coding problem entails identifying which aspects of neuronal activity subserve informational functions and phenomenal distinctions (which aspects of neural activity constitute differences that make a difference for mental functions and subjective experiences). For perception, the neural coding problem entails identifying which aspects of neuronal activity are related to different perceptual attributes (e.g. for a musical note, its loudness, duration, pitch attributes, timbral qualities, and apparent spatial location).

A working assumption is that in order for neural signals to be propagated and regenerated within the system their constituent patterns of neural activity must be organized in a form that is meaningful to the system, i.e. it must be in the form of a recognizable *neural code* (Perkell and Bullock 1968; Cariani 1995a, 2001c).

Neural codes can be divided into channel codes and temporal codes, depending on how informational distinctions are represented within the system, i.e. which aspects of neuronal activity convey information. Channel codes depend on which neurons are responding, and rate-channel codes depend on profiles of firing rates across neurons, i.e. which neurons are firing how rapidly. Temporal codes, on the other hand, depend on different patternings of spike timings, rather than which particular neurons are producing them. Figure 4A shows some examples of simple and complex temporal codes.

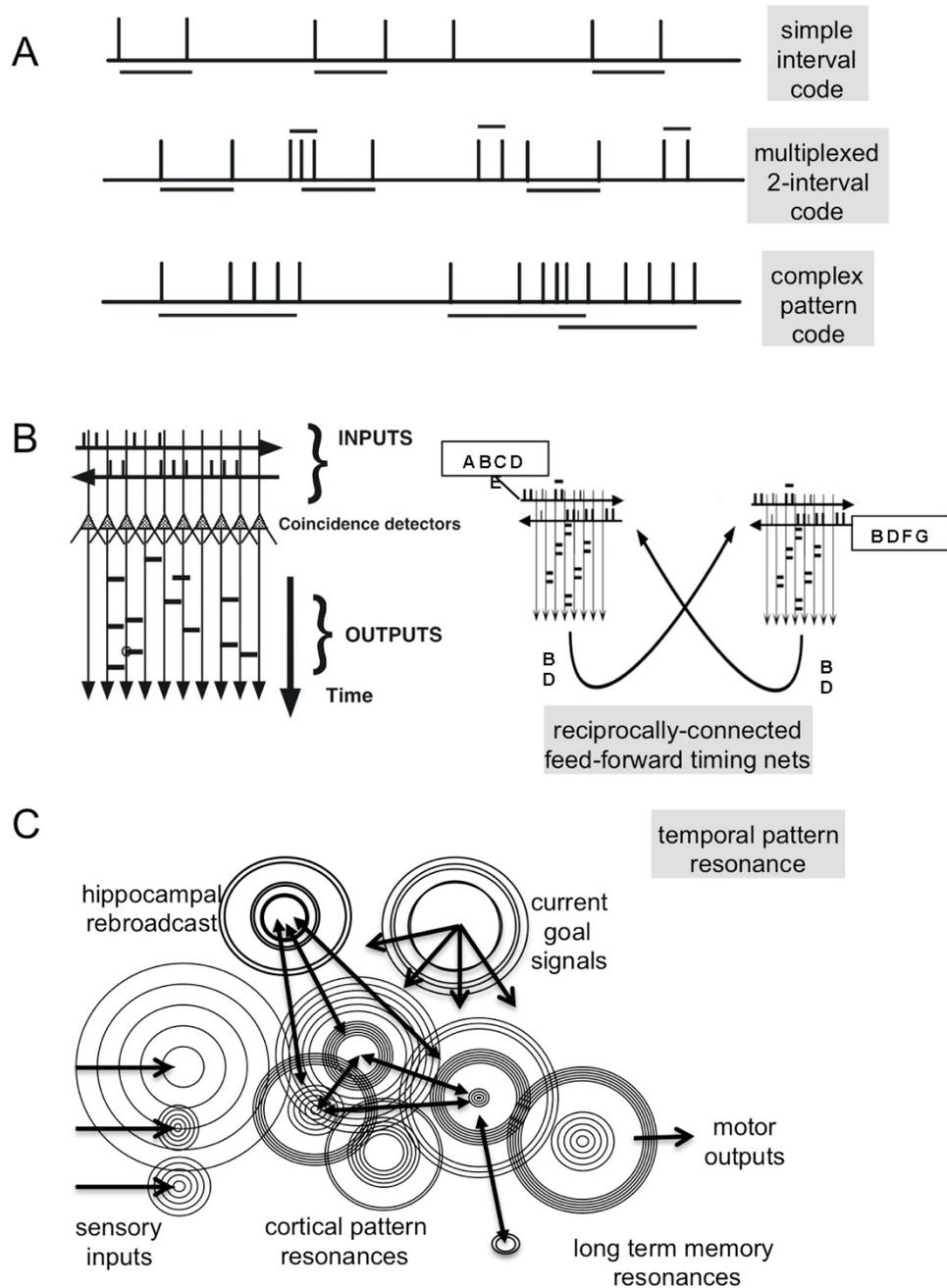


Figure 4. Temporal codes, neural timing nets, and signal dynamics. A. Examples of simple and more complex temporal codes. Complex temporal codes permit signal multiplexing – multiple types of distinctions to be conveyed over the same transmission lines. B. Neural timing nets and signal dynamics. When signals in two reciprocally-connected regions share the similar feature-related subpatterns, they mutually reinforce each other, such that the common signal-dimensions build up. C. Interference and reinforcement dynamics of neural signals. Broadcast of temporally coded signals. Neural assemblies tuned to particular combinations of temporal patterns emit characteristic tag patterns that indicate the recognition of that particular feature-combination. From Cariani (2015).

Temporal pulse codes have a number of advantages vis-a-vis rate-channel codes. First, signals are no longer tied to particular transmission lines or paths (liberation of signals from wires permits broadcast strategies of coordination). Simple signals can be combined (multiplexed) to form vectorial signals that can be multiplexed, i.e. transmitted concurrently over the same lines.

Temporal codes. Temporal codes can be found in many diverse places in the nervous system (Cariani 1995b, 2001c; Mountcastle 1967). I have investigated the neural coding of pitch, mainly in early auditory stations, and the evidence for temporal coding of pitch at that level is quite strong. Given the statistics of neural spike timing patterns (interspike intervals) it is possible to predict with a high degree of accuracy and reliability which pitches will be heard under normal circumstances (Cariani and Delgutte 1996; Cariani 1999). Random, uncorrelated neural activity does not normally appear to contribute to specific perceptual qualities or we would constantly experience the flicker, roar, smell, and taste of spontaneous activity in our brains.

All distinctions that our brains make and of which we are consciously aware are presumed to be conveyed, processed, stored, and regenerated in the form of properly structured neural codes. Further, it is hypothesized that all attributes associated with sensations, thoughts, feelings, motivations, and memories are encoded in a single framework, a *lingua franca* of the brain that is analogous to that of the cell, the genetic code.

Brain function through signal dynamics. The temporal theory of brain function holds that the brain is a vast network of delay loops in which neural signals in the form of correlated temporal patterns of spikes are actively regenerated through neural pattern-amplification processes. The functional states of the brain involve recurrent activity in which signals persist over time (tens to hundreds of milliseconds or more). Those sets of neural signals that are actively circulating at any given time determine the contents of awareness, whereas the ability of the system to actively regenerate signals determines the state of consciousness (e.g. waking vs. anesthetized). Which circuits of the brain can support active regeneration determine the particular state of consciousness. For example, the waking, alert state is associated with regeneration in both local and global networks, whereas sleep is associated with regeneration in only local networks. Trance, hypnosis, meditation, and other kinds of states may be associated with particular patterns of signal regeneration in subsets of circuits.

The crucial difference between those waking states that support conscious awareness and anesthesia- and seizure-produced states that do not is due entirely to the differences in the organization of neural activity. The main effect of general anesthetic agents in abolishing awareness may involve disruption of neuronal temporal response properties that are critical for signal regeneration, in effect scrambling temporally coded signals rather than suppressing gross levels of neural activity (Cariani 1997; Cariani 2000a).

The proposed theory is very similar to neural local and global workspace theories of the neural basis of consciousness in that these also hold that the contents of our awareness are determined by patterns of activity in global and/or local recurrent networks. The reason that recurrent activity is critical is due to the self-regeneration of neural signals – closing the loop allows signals to amplify themselves and therefore to persist within the network.

Although the temporal theory has some commonalities with Varela and Thompson's theory that relies on dynamics of nonlinear oscillators (Thompson and Varela 2001), it differs from that theory in that the temporal theory relies on asynchronous temporal spike patterns rather than on neuronal synchronies and oscillations (the neural correlates of perceptual features were never clearly spelled out in their theory).

The temporal theory also assumes that there is a strong dimensional structure to perception (not unlike the structuralism of Wundt, Titchener, and Boring (Boring 1942)) that is a reflection of a the dimensional structure of neural coding space (Köhler's concept of psychoneural isomorphism between percepts and underlying neural processes).⁴

The temporal theory comes out of temporal models of pitch (Cariani 1999; Cariani and Delgutte 1996), which explain a wide array of pitch percepts in terms of patterns of spikes in neuronal ensembles – temporal codes based on interspike interval patterns (Fig. 4A top spike train). The temporal theory differs from connectionist models in that the neural signals are temporal patterns of spikes present in populations of neurons (spike correlation patterns, irrespective of which particular neurons are involved) rather than patterns of average firing rates amongst neurons (which neurons fire how often). Arguably, purely connectionist architectures, because of their need to highly regulate neural transmission paths and connection-weights, are inadequate to account for the flexibility of mental processes.

As an alternative to connectionism, I have proposed neural timing nets (Fig. 4) that consist of arrays of delay lines and spike-timing-facilitated coincidence elements that would implement reverberating short-term memory circuits and pattern analysis (Cariani 2001a; Cariani 2002; Cariani 2004). The neural timing networks are "signal-centric" in that the expressive combinatoric potential of the system relies on complex multidimensional signals and relatively simple neural substrates. This contrasts with the anatomical, "connection-centric" connectionist schemes which rely on simple scalar signals (running firing rates) and very complex patterns of synaptic connections and efficacies. Connectionist networks operate through summation and thresholding of weighted scalar signals, whereas timing nets carry out "delay path computations."

Complex temporal codes potentially support multidimensional representations, signal multiplexing, and broadcast-based coordinations (Cariani 2017, 2015a). Complex temporal codes also potentially support signal dynamics, in which complex temporal patterns interact to mutually reinforce those subpatterns that are common to them. It is then possible to envision a bottom-up dynamics of neural signals that obviates the need for the symbols-and-rules schemas that are typical of computationalist models of cognitive processes.

In the neural timing theory, the regeneration of signals in neural delay loops creates self-sustaining patterns of neural activity. Signal regeneration is needed to create sustained functional brain states that can switch behavior. The ability to regenerate signals within global loops is the necessary and sufficient condition for conscious awareness in the waking state. The particular signals that are regenerated in global loops at any given time determine the contents of awareness. This hypothesis is similar in many respects to current theories of the neural requisites for conscious awareness that rely on re-entrant neuronal activity in local and global circuits (Dehaene 2014; Lamme 2006).

An analogy can be made with autopoiesis, which entails networks of material components and relations that interact to regenerate the constituent components and relations. In the neural case, rather than material components, distinctions in the form of internally-distinguishable pulse patterns (neural signals) are regenerated (Cariani 2000b). Such self-production networks realize organizational closure by creating an inside and an outside. The inside is an internal realm of autonomous circular-causal control by the self-production network and the outside is an external realm that is not under the control of the self-production network.

⁴ Neural dynamical systems typically have much, much higher degrees of freedom, so to effectively model the lower dimensional structure of perception, with its stimulus equivalences and relative invariances under transformation, it is necessary to find lower dimensional descriptions of neural dynamics.

The theory is still rudimentary and its details are still being worked out, but what it means is that there are alternatives to both the symbols-and-rules computations of symbolic AI and symbolic cognitive science and the standard connectionist neural networks.

Similarities with other theories. The theory does share many assumptions with connectionist theories of brain function and consciousness (Rose 2006). Both connectionist and neural timing theories assume that our immediate awareness is mediated by short-term memory mechanisms (sensory memory, working memory), which is labile such that it is subject to disruption by sleep, anesthesia, attentional changes, and additional stimuli. Long-term memory is mediated by a quasi-permanent storage and retrieval process that is highly resilient to these aforementioned disruptions. Both theories hold that the contents of our awareness depend on the neural activity patterns that are actively regenerated in reverberating short-term memory loops. Connectionist theories emphasize ensembles of neurons that maintain higher firing rates through short-term synaptic facilitations, whereas the temporal theory emphasizes the active maintenance of temporal patterns in reverberating delay loops through spike timing dependent facilitations.

Differences with complexity-based theories. The theory diverges from those that postulate a necessary threshold degree of informational complexity for awareness (Tononi 2005). In the theory proposed here organizational closure (through signal regeneration) rather than informational complexity per se is the critical organizational substrate.

Differences with strong enactivism. This proposed neuropsychological and neurophenomenal theory is embodied and supports mental functions that are extended into the environment (e.g. extended cognition). It is weakly enactivist in that it holds that externalized action is the central *raison d'être* for most mental functions. However, it departs from strong enactivism on several fronts: 1) there are other modes of activity that are inwardly focused such that action is not the absolute end-all and be-all of existence⁵, 2) conscious awareness is a product of activity in the brain (consciousness and meaning are in the head), and 3) although perception and other mental operations are influenced by action, strong motor theories of perception (i.e. no perception without motor system activation) are easily falsified (Hickok 2014). The relation between processes of action and those of perception, cognition, affect, conation, and memory appears to be more bi-directional and reciprocal in nature than would be assumed by strong motor theories. In my view, each process has some relative autonomy vis-a-vis the others.

Some of the disagreements may also be more rhetorical than substantive, and stem from redefinitions of terms such as consciousness, cognition, information, representation, and computation. In many of the talks about the philosophy of enactivism at the conference, a

⁵ Action isn't everything we do or experience. Externalized action is not the only activity that mind and brain engage in. In listing different general modes for audition, I count an *ecological mode*, in which we listen in order to gain information about our environments for purposes of orientation and action, a *communicative mode*, in which we listen to speech in order to discern a message, an *aesthetic mode*, in which we listen (usually to music) in order to modulate our own internal psychological states in desirable ways, and an *annoyance mode*, in which we are forced to listen to sounds that interfere with our desires (goals). Orientation is not action *per se*. Understanding speech is not itself action. In the case of listening to music in the aesthetic mode, the purpose does not involve externalized action on the world, but an internal change in the contents of awareness. Many forms of meditation and thoughtful reflection are similarly inwardly-directed. In contrast to strong-enactivist perspectives that mandate that everything we experience must necessarily be related to externalized action, these counterexamples suggest that while the coordination of action may be the most important of all mental activities, certainly in terms of survival and reproduction, there are nevertheless realms of mental activity and experience that are not themselves directly related to it. Action (overt behavior) is not all that there is to brain, mind, and experience.

dichotomy of perspectives was presented that pitted symbolic AI and computationalist cognitivism against dynamicist-enactivist frameworks, without any third or fourth alternatives. Discussions of these definitions and perspectives are sometimes much more nuanced and flexible in the enactivist literature (Thompson 2007).

I have agreed all along with many of the deep criticisms leveled at the platonic ontology and realist epistemology that animates computationalism (Cariani 1989). The computationalist ideology marginalizes interactions with the environment (perception and action) and precludes even the possibility of the emergence of fundamental novelty (Cariani 1992; Cariani 2012). However, in formulating alternatives, it is critical to avoid false dichotomies, because there are almost invariably multiple perspectives that incorporate, subsume, and sometimes transcend the presumed dichotomies. It is absolutely necessary to get beyond both symbols and rule-systems and low-level dynamics if we are to understand minds and brains work (Cariani 2001b). In this respect my approach, at the level of interactions of analog temporally-patterned neural signals, is similar to the middle-level representations of Prinz (Prinz 2012).

IV. Temporal commonalities of perception and action

Traditional theories of brain function are largely atemporal in that they assume that perception and action are mediated by average rates of neural firing. However, there is strong evidence that many percepts depend on stimulus-driven temporal patternings of spikes. Here sensory stimuli impress their temporal structure on the firings of neurons, on both microtemporal scales (e.g. pitch, timbre) and on macrotemporal scales (rhythm).

Rhythm may well be directly temporally-coded. Rhythmic patterns consisting of discrete sensory events separated by 100 ms or more produce brisk event-onset responses that reproduce the temporal structure of the rhythms in sensory pathways up to and including the cortex. It has been known for quite a long time that such rhythmic patterns can be observed in cortical gross potentials (auditory evoked potentials, event-related potentials). For the most part, rhythm has been modeled in terms of clocks, modulation-tuned neurons, and nonlinear oscillators rather than as a direct temporal code. Recently temporal correlates of both experienced and imagined rhythms have been found (Nozaradan 2014; Nozaradan et al. 2013; Nozaradan et al. 2011). These would be similar to the temporal patterns that would be generated in motor regions in producing the same rhythms.

Recently, experiments with passive haptic learning have demonstrated that presenting temporal patterns of experienced pianists to the fingertips of untrained subjects greatly accelerates the learning of piano pieces (Starner 2013).

Such temporal codes produce iconic internal neural representations that import the correlation structure of the environment into the nervous system. Likewise, on the motor side, coordinated action requires precise temporal patterning (relative timing) of muscular activations. When muscles contract, stretch receptors and mechanoreceptive afferents produce spikes that are phase-locked to the resulting movements. These temporal patternings of body movements are in turn fed back into the brain such that the brain is continually bathed in the temporal structure of actions. The ensuing action on the environment, such as drumming the rhythmic pattern, creates still other temporal feedbacks in other sense modalities (vision, touch, hearing).⁶

⁶ It is also conceivable that our emotions are conveyed via characteristic temporal pattern codes (Clynes 1977), such that music and dance evoke emotional feelings by directly, albeit weakly, creating internal temporal patterns that mimic those codes. There exist pervasive parallels between music, movement, and emotion (Sievers et al. 2013). This

Thus, there may be a common neural language that underlies the temporal aspects of perception and action, such that perception and action can mutually inform each other in a direct manner. Hearing a musical rhythm provides a temporal scaffold for timing movements, and the timing of movements reinforces the perceived rhythmic structure of the music.

V. Acknowledgement

I would like to thank the Office for Faculty Development at the Boston Conservatory @ Berklee College of Music for generous travel support that enabled me to attend this meeting.

VI. References

- Ackoff RL, Emery FE (1972) *On Purposeful Systems*. Aldine-Atherton, Chicago,
- Ashby WR (1956) *An Introduction to Cybernetics*. Chapman and Hall, London
- Boring EG (1933) *The Physical Dimensions of Consciousness*. Dover, New York
- Boring EG (1942) *Sensation and Perception in the History of Experimental Psychology*. Appleton-Century-Crofts, New York
- Cariani P (1989) *On the Design of Devices with Emergent Semantic Functions*. Ph.D., State University of New York at Binghamton, Binghamton, New York
- Cariani P (1992) Emergence and artificial life. In: Langton CG, Taylor C, Farmer JD, Rasmussen S (eds) *Artificial Life II. Volume X*, Santa Fe Institute Studies in the Science of Complexity. Addison-Wesley, Redwood City, CA, pp 775-798
- Cariani P (1995a) As if time really mattered: temporal strategies for neural coding of sensory information. Reprinted in: K Pribram, ed *Origins: Brain and Self-Organization*, Hillsdale, NJ: Lawrence Erlbaum, 1994; 208-252 12 (1-2):161-229 (Reprinted in: K Pribram, ed. *Origins: Brain and Self-Organization*, Hillsdale, NJ: Lawrence Erlbaum, 1994; 1208-1252.)
- Cariani P (1995b) As if time really mattered: temporal strategies for neural coding of sensory information. In: Pribram K (ed) *Origins: Brain and Self-Organization*. Lawrence Erlbaum, Hillsdale, NJ, pp 208-252
- Cariani P (1997) Consciousness and the organization of neural processes: commentary on John et al. *Conscious Cogn* 6 (1):56-64
- Cariani P (1999) Temporal coding of periodicity pitch in the auditory system: an overview. *Neural Plasticity* 6 (4):147-172
- Cariani P (2000a) Anesthesia, neural information processing, and conscious awareness. [Commentary on Alkire et al]. *Consciousness and Cognition* 9 (3):387-395
- Cariani P (2000b) Regenerative process in life and mind. *Ann N Y Acad Sci* 901:26-34
- Cariani P (2001a) Neural timing nets. *Neural Networks* 14 (6-7):737-753
- Cariani P (2001b) Symbols and dynamics in the brain. *Biosystems* 60 (1-3):59-83
- Cariani P (2001c) Temporal coding of sensory information in the brain. *Acoust Sci & Tech* 22 (2):77-84
- Cariani P (2002) Temporal codes, timing nets, and music perception. *J New Music Res* 30 (2):107-136

would be a means by which the *form* of music and dance could itself be emotional (Langer 1942), and it provides a possible explanation for why these temporal art forms have the manifold expressive emotional powers that they do.

- Cariani P (2004) Temporal codes and computations for sensory representation and scene analysis. *IEEE Transactions on Neural Networks, Special Issue on Temporal Coding for Neural Information Processing* 15 (5):1100-1111
- Cariani P (2011) The semiotics of cybernetic percept-action systems. *International Journal of Signs and Semiotic Systems* 1 (1):1-17
- Cariani P (2012) Creating new primitives in minds and machines. In: McCormack J, D'Inverno M (eds) *Computers and Creativity*. Springer, New York, pp 395-430
- Cariani P (2015a) Outline of a cybernetic theory of brain function based on neural timing nets. *Kybernetes* 44 (8/9):1219-1232
- Cariani P (2015b) Sign functions in natural and artificial systems. In: Trifonas PP (ed) *International Handbook of Semiotics*. Springer, Dordrecht, pp 917-950
- Cariani P (2017) Temporal memory traces as anticipatory mechanisms. In: Nadin M (ed) *Anticipation and Medicine*. Springer, Switzerland, pp 105-136
- Cariani PA, 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
- Clynes M (1977) *Sentics : the touch of emotions*. 1st edn. Anchor Press, Garden City, N.Y.
- Craik KJ (1966) *The Nature of Psychology. A Selection of Papers, Essays, and Other Writings by Kenneth J. Craik*, Edited by Stephen L. Sherwood. Cambridge University Press, Cambridge
- de Latil P (1956) *Thinking by Machine*. Houghton Mifflin, Boston
- Dehaene S (2014) *Consciousness and the brain : deciphering how the brain codes our thoughts*. Viking, New York, New York
- DeYoung CG (2015) Cybernetic Big Five Theory. *Journal of Research in Personality* 56:33-58
- Fessard AE (1954) Mechanisms of nervous integration and conscious experience. In: Adrian ED, Bremer F, Jasper HH (eds) *Brain Mechanisms and Consciousness*. Charles C. Thomas, Springfield, IL, pp 200-236
- Graham DW (1987) *Aristotle's Two Systems*. Oxford University Press, New York
- Hickok G (2014) *The myth of mirror neurons : the real neuroscience of communication and cognition*. First edition. edn. W. W. Norton & Company, New York
- Jones MR (1976) Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychol Rev* 83 (5):323-355
- Kazansky AB (2015) Agental anticipation in the central nervous system. In: Nadin M (ed) *Anticipation: Learning from the Past: The Russian/Soviet Contributions to the Science of Anticipation*, vol 25. Cognitive Systems Monographs. Springer, New York, pp 108-117
- Kilmer W, McCulloch WS (1969) The Reticular Formation Command and Control System. In: McCulloch R (ed) *The Collected Works of Warren McCulloch*, vol IV. Intersystems Publications, 1989, Salinas, CA, pp 1322-1332
- Koch C (2004) *The Quest for Consciousness : a Neurobiological Approach*. Roberts & Co., Denver
- Köhler W (1947) *Gestalt Psychology*. Liverwright, New York
- Lamme VA (2006) Towards a true neural stance on consciousness. *Trends Cogn Sci* 10 (11):494-501
- Langer SKK (1942) *Philosophy in a new key; a study in the symbolism of reason, rite and art*. Harvard University Press, Cambridge, Mass.,
- Maturana HR, Varela F (1980) *Autopoiesis and Cognition: The Realization of the Living*. D. Reidel, Dordrecht
- McCulloch WS (1946) A heterarchy of values determined by the topology of nervous nets. In: McCulloch WS (ed) *Embodiments of Mind*. MIT Press, Cambridge, Mass., pp 40-45. Originally appeared in *Bull. Math. Biophys.* 47(42): 89-93

- Modrak DK (1987) Aristotle: The Power of Perception. University of Chicago, Chicago
- Mountcastle V (1967) The problem of sensing and the neural coding of sensory events. In: Quarten GC, Melnechuk, T., and Schmitt, F.O. (ed) *The Neurosciences: A Study Program*. Rockefeller University Press, New York,
- Nozaradan S (2014) Exploring how musical rhythm entrains brain activity with electroencephalogram frequency-tagging. *Philos Trans R Soc Lond B Biol Sci* 369 (1658). doi:10.1098/rstb.2013.0393
- Nozaradan S, Peretz I, Missal M, Mouraux A (2011) Tagging the neuronal entrainment to beat and meter. *J Neurosci* 31 (28):10234-10240. doi:10.1523/JNEUROSCI.0411-11.2011
- Nozaradan S, Zerouali Y, Peretz I, Mouraux A (2013) Capturing with EEG the Neural Entrainment and Coupling Underlying Sensorimotor Synchronization to the Beat. *Cereb Cortex*. doi:10.1093/cercor/bht261
- Pattee HH, Raczaszek-Leonardi J (2012) *Laws, language and life* Howard Pattee's classic papers on the physics of symbols with contemporary commentary by Howard Pattee and Joanna Raczaszek-Leonardi. *Biosemiotics v 7*. Springer, Dordrecht ; New York
- Perkell DH, Bullock TH (1968) Neural Coding. *Neurosciences Research Program Bulletin* 6 (3):221-348
- Powers W (1973) *Behavior: The Control of Perception*. Aldine, New York
- Prinz JJ (2012) *The conscious brain : how attention engenders experience*. Philosophy of mind. Oxford University Press, Oxford ; New York
- Rose D (2006) *Consciousness : philosophical, psychological and neural theories*. Oxford University Press, Oxford ; New York. doi:9780198792949
- Rosenblueth A, Wiener N (1950) Purposeful and Non-Purposeful Behavior. *Philos. of Science* 17 (4):318-326
- Rosenblueth A, Wiener N, Bigelow J (1943) Behavior, purpose and teleology. *Philos. of Science* 10 (S):18-24
- Sievers B, Polansky L, Casey M, Wheatley T (2013) Music and movement share a dynamic structure that supports universal expressions of emotion. *Proc. Natl. Acad. Sci. USA* 110 (1):70-75. doi:10.1073/pnas.1209023110
- Starner T (2013) *Wearable computing: Thad Starner at TEDxSalford 2013*.
- Thompson E (2007) *Mind in life : biology, phenomenology, and the sciences of mind*. Belknap Press of Harvard University Press, Cambridge, Mass.
- Thompson E, Varela FJ (2001) Radical embodiment: neural dynamics and consciousness. *Trends Cogn Sci* 5 (10):418-425. doi:S1364-6613(00)01750-2 [pii]
- Tononi G (2005) Consciousness, information integration, and the brain. *Prog Brain Res* 150:109-126
- Troland LT (1929) *The Principles of Psychophysiology, Vols I-III*. D. Van Nostrand, New York
- Uexküll Jv (1926) *Theoretical Biology*. Harcourt, Brace & Co, New York
- Varela F (1996) Neurophenomenology: A Methodological remedy to the hard problem. *Journal of Consciousness Studies* 3 (4):330-349
- Zueva EY, Zuev KB (2015) The concept of dominance by A.A. Ukhtomsky and Anticipation. In: Nadin M (ed) *Anticipation: Learning from the Past: The Russian/Soviet Contributions to the Science of Anticipation, vol 25*. Cognitive Systems Monographs. Springer, New York, pp 13-35