

A Cortical Network Model of Cognitive Attentional Streams, Rhythmic Expectation, and Auditory Stream Segregation

Bill Baird

Dept Mathematics, U.C. Berkeley,
Berkeley, Ca. 94720. baird@math.berkeley.edu

We have developed a neural network architecture that implements a theory of attention, learning, and communication based on adaptive synchronization of 5-15 Hz and 30-80 Hz oscillations between cortical areas [Baird et al., 1994]. Here we present a specific model of rhythmic expectancy and the interaction of higher-order and primary cortical levels of processing which accounts for the results of psychological experiments of Jones [Jones et al., 1981] showing that auditory stream segregation depends on the rhythmic structure of inputs. Further references not cited here may be found in these papers.

Using dynamical systems theory, the architecture is constructed from recurrently interconnected oscillatory associative memory modules that model hypercolumns of associational and higher-order sensory and motor cortical areas. The modules learn interconnection weights that cause the system to evolve under a 10 Hz clocked sensory/motor processing cycle through a sequence of transitions of synchronized 40 Hz oscillatory attractors within the modules. In the brain, we hypothesize these cycles to be adaptively controlled by septal and thalamic pacemakers which alter excitability of hippocampal and neocortical tissue through nonspecific biasing currents that appear as the cognitive and sensory evoked potentials of the EEG. The cycles "quantize time" and form the basis of derived rhythms with periods up to 1.5 seconds that entrain to each other in motor coordination and to external rhythms in speech and music perception.

The architecture employs selective "attentional" control of the synchronization of the 30-80 Hz oscillations between modules to direct the flow of communication and computation in learning to recognize and generate sequences. The 30-80 Hz attractor amplitude patterns code the information content of a cortical area, whereas phase and frequency are used to "software" the network, since only the synchronized areas communicate by exchanging amplitude information. The system works like a broadcast network where the unavoidable crosstalk to all areas from previous learned connections is overcome by frequency coding to allow the operation of attentional communication only between selected areas relevant to the task of the moment. The behavior of the time traces in different modules of the architecture models the temporary appearance and switching of the synchronization of 5-15 and 30-80 Hz oscillations between cortical areas that is observed during sensory/motor tasks in monkeys and humans.

The model architecture illustrates the notion that synchronization not only preattentively "binds" the features of inputs in primary sensory cortex into

"objects", but further binds the activity of an attended object to oscillatory activity in associational and higher-order sensory and motor cortical areas to create an evolving attentional network of intercommunicating cortical areas that directs behavior. This is a model of "attended activity" as that subset which has been included in the selectively attended processing of the moment by synchronization to this network. This involves both a spatial binding of activity and a binding of the last to the next step of a sequence. Only inputs which are synchronized to the internal oscillatory activity of a module can effect the proper learned *transitions* of attractors within it.

The phenomenon of "streaming" in audition most exemplifies this notion of *sequence binding*. There successive events of a sound source are bound together into a sequence object or "stream" and segregated from other sources such that one can pay attention only to one sound source at a time - as in the "cocktail party" phenomenon. "Cognitive streams" are in evidence when two stories are told in alternating segments and listeners are unable to recall the relative order of events between them. We view the model attentional network as a stream because the synchronized modules within it are cycling through sequences of attractors at the 10 Hz rate. The feedback from higher-order to primary cortical areas allows top down voluntary control to switch the **primary attention stream** or "searchlight" from one source preattentively bound in primary cortex to another source separately bound at a nearby frequency.

There is evidence from studies of motor and perceptual tasks that motor and perceptual behavior is organized by neural rhythms with periods in the range of 100 - 1500 milliseconds, and that entrainment of these to external rhythms in speech and other forms of communication, and to internal rhythms in motor coordination is essential to effective human performance [Jones et al., 1981]. In this view, just as two cortical areas must synchronize to communicate, so must two nervous systems. Work with slowed film suggests that both speaker and listeners show synchronization of body movements to the beat of the roughly 10 Hz rate of phoneme emission. Steady stimulation at either 10 or 40 Hz in audition, vision, or somatosensation causes entrainment of 10 or 40 Hz activity in those areas. 20 Hz stimulation entrains the 40 Hz activity at a 1:2 resonance ratio and has been used medically as a diagnostic for proper function of cortical areas.

Rhythms related to attentional expectation have been found in the auditory EEG. In experiments where the arrival time of a target stimulus is reg-

ular enough to be learned by an experimental subject, it has been shown that the 10 Hz activity in advance of the stimulus becomes phase locked to that expected arrival time. The same has been shown for hippocampal theta in rats, which is also found to be entrained to the speed of locomotion of a rat. Rats palpate an object with their paws at 10 Hz and will entrain this "active touch" to an object vibrating near this rate. The "mismatch negativity" of the auditory evoked potential is a reliable indicator of the action of a physiological expectancy system. Changes in the onset or offset time of a repetitive stimulus produce an increase in the N200 component in primary auditory cortex. This is physiological evidence for Jones' theory postulating a rhythmic expectancy system.

Jones [Jones et al., 1981] replicated and altered a classic streaming experiment of Bregman and Rudnick, and found that their result depended on a specific choice of the rhythm of presentation. The experiment required human subjects to determine the order of presentation of a pair of high target tones AB or BA of slightly different frequencies. Also presented before and after the target tones were a series of identical much lower frequency tones called the capture tones CCC and two identical tones of intermediate frequency before and after the target tones called the flanking tones F - CCCFABFCCC.

Target order determination performance deteriorated as the flanking tones were brought near to the target tones in frequency, implying that the flanking tones were captured by that stream and obscured the target tones. Otherwise the flanking tones were captured by the background capture tone stream, leaving the target tones to stand out by themselves in the attended stream. Jones noted that the flanking tones and the capture stream were presented at a stimulus onset rate of one per 240 ms and the targets appeared at 80 ms intervals. In her experiments, when the target tones were given a tempo in common with the capture and flanking tones, no effect of the frequency of flanking tones appeared. This suggested that rhythmic distinction of targets and distractors was paramount over frequency in allowing selective attention to segregate out the target stream.

To implement Jones's theory in the model and account for her data, subsets of the oscillatory modules are dedicated to form a rhythmic temporal coordinate frame or time base of nested periodicities in divisions of the thalamic 10 Hz base clock rate between 10 to .5 Hz. Each periodicity is created by an associative memory module that has been specialized to act stereotypically as a counter or shift register by repeatedly cycling through all its attractors at the rate of one for each time step of the clock. Its overall cycle time is therefore determined by the number of attractors. Only one step of the cycle can send output to primary cortex - the one with the largest weight from receiving the most matches to incoming stimuli. Each clock derived in this manner from the thalamic base clock will therefore phase reset itself to get the best match to incoming rhythms. The match is further refined by phase adjustment of the base clock

itself.

The modules of the time base send their internal 30-80 Hz activity to primary auditory cortex in 100msec bursts at their different rhythmic rates through fast adapting connections that continually attempt to match incoming stimulus patterns. These weights effectively compute a rough low frequency discrete Fourier transform over a sliding window of 1.5 seconds. Those temporal stimulus patterns which meet these established rhythmic expectancy signals in time are pulled into synchrony with the 30-80 Hz activity of the time base to form a "stream" of events defined by the common synchrony of each with the time base. Stimuli up to a second apart may thus be bound into the same stream without requiring activity to sit undisturbed in the auditory buffer for that long. Segregation of streams in auditory cortex by frequency, intensity, timbre, and other spatially distributed features is done by frequency segregation of the activity within the 30-80 Hz range. In accordance with Jones' theory, voluntary top-down attention can probe input at different hierarchical levels of periodicity by selectively synchronizing with a particular cortical patch in the time base set at the particular 40 Hz frequency of the primary attention stream. Then the "searchlight" into primary cortex is synchronized with and reading in activity occurring at that rhythm. This is the temporal analog of the body centered spatial coordinate frame and multiscale covert attention window system in vision. Here the body centered temporal coordinates of the internal time base orient by entrainment to the external rhythm, and the window of covert temporal attention can then select a level of the multiscale temporal frame.

Our explanation of Jones' result is that the standard target tones first prime the dynamic attention system, then the slow capture tones establish a background stream with a rhythmic expectancy that is later violated by the fast target tones. These are driven into a separate stream which allows the top down primary attention stream, already primed to look for that rhythm, to synchronize and read in the target tones for order determination. In the absence of a rhythmic distinction for the target tones, their pitch difference alone is insufficient to drive the formation of a separate stream, and the mechanism of dynamic attention cannot help discriminate the targets.

References

- [Baird et al., 1994] Baird, B., Troyer, T., and Eeckman, F. H. (1994). Attention as selective synchronization of oscillating cortical sensory and motor associative memories. In Eeckman, F. H., editor, *Neural Systems Analysis and Modeling*, Norwell, Ma. Kluwer. in press.
- [Jones et al., 1981] Jones, M., Kidd, G., and Wetzel, R. (1981). Evidence for rhythmic attention. *Journal of Experimental Psychology: Human Perception and Performance*, 7:1059-1073.