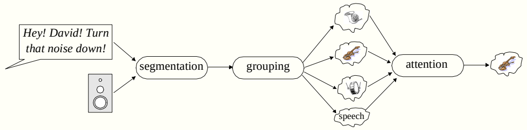
A Neural Oscillator Model of Binaural Auditory Selective Attention

Stuart N. Wrigley and Guy J. Brown

Department of Computer Science, University of Sheffield, UK

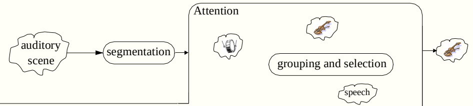
s.wrigley@dcs.shef.ac.uk http://www.dcs.shef.ac.uk/~stu g.brown@dcs.shef.ac.uk http://www.dcs.shef.ac.uk/~guy

'Classical' Auditory Scene Analysis



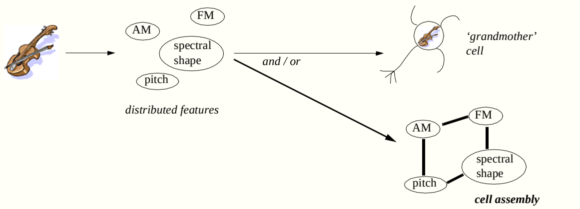
Attention in ASA

- It is now believed that attention plays a much more prominent role in ASA - attention does more than simply select a single stream (Carlyon *et al.*, 2001).



- *Exogenous* attention groups acoustic elements
- *Endogenous* mechanisms performs stream segregation and selection.

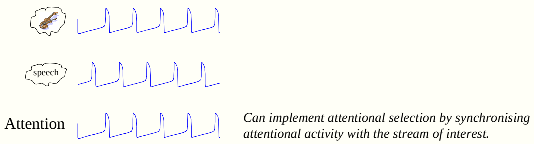
Binding problem



- Additionally, only attended streams are encoded into memory and perceived (Moore and Egeth, 1997) - attention controls which percepts are to be encoded into memory.

Oscillatory Correlation Framework

- The oscillatory correlation theory (Wang, 1996) suggests that neural oscillations encode auditory grouping - i.e. a solution to the binding problem (see also von der Malsburg, 1981).

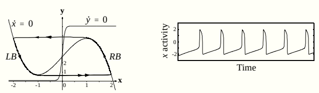


Relaxation Oscillator (Wang, 1996)

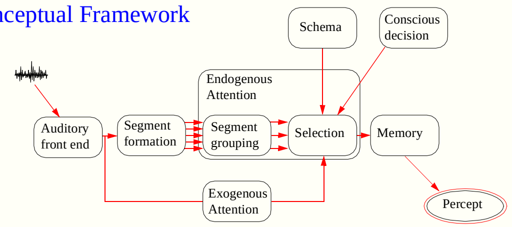
- Reciprocally connected excitatory unit and inhibitory unit whose activities are represented by x and y.
- Conceptually, the oscillator represents mean activity of a population of neurons. An alternative concept is that of the behaviour of a neuron's membrane potential and ion channels.

$$\dot{x} = 3x - x^3 + 2 - y + I_o$$

$$\dot{y} = \varepsilon\left[\gamma\left(1 + \tanh\frac{x}{\beta}\right) - y\right]$$



Conceptual Framework

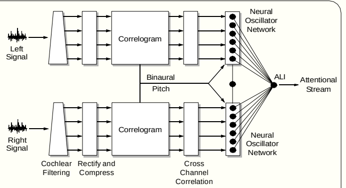


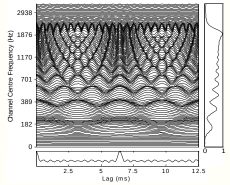
Abstract

A model of auditory grouping is described in which auditory attention plays a key role. The model is based upon an oscillatory correlation framework, in which neural oscillators representing a single perceptual stream are synchronised, and are desynchronised from oscillators representing other streams. The model suggests a mechanism by which attention can be directed to the high or low tones in a repeating sequence of tones with alternating frequencies. In addition, it simulates the perceptual segregation of a mistuned harmonic from a complex tone.

Model Overview

- Cochlear filtering is performed by 128 gammatone filters with centre frequencies equally spaced on the ERB scale between 50 Hz and 3.5 kHz.
- Auditory nerve firing rate is approximated by half-wave rectifying and square root compressing the output of each filter.





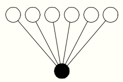
- Pitch information is extracted by computing the autocorrelation of the activity in each channel to form a correlogram (Brown and Cooke, 1994).
- The correlogram is used to identify formant and harmonic regions by locating areas which exhibit similar patterns of periodicity (Wang and Brown, 1999). This is achieved by computing the cross-channel correlation of the correlogram.
- Each cross-frequency area exhibiting similar patterns of periodicity is classed as a *segment*.

Neural Oscillator Network

- The network consists of 128 oscillators per ear.
- Segment formation: excitatory connections promote synchronisation of oscillator groups.
- Segments are (primitively) grouped on the basis of harmonicity and stimulus 'age' (old-plus-new heuristic) by placing local excitatory links between them.
- Input I_o to each oscillator is combination of raw input (I_r - based on segments), global inhibition and network activity.

$$I_o = I_r - W_z S(z, \theta_z) + \sum_{k \neq i} W_{ik} S(x_k, \theta_x)$$

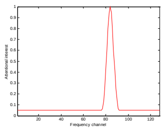
where x_k is oscillator activity at channel k and S is a 'squash' function. W_{ik} is the connection strength between oscillators i and k and W_z is the weight of inhibition from the global inhibitor z.



- Each oscillator in the network feeds excitatory input to the global inhibitor (a leaky integrator). The global inhibitor, in turn, feeds inhibitory input back to each oscillator.
- The global inhibitor ensures that only one block of synchronised oscillators can be active at any one time. Hence, separate blocks of synchronised oscillators (corresponding to 'elements' in ASA) arise through the action of local excitation and global inhibition.

Attentional Leaky Integrator

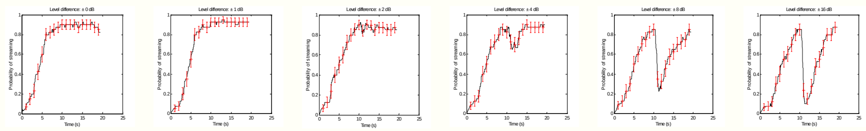
- The strength of the oscillator to ALI connection is determined by 'conscious' attentional interest which is directed toward a particular frequency region.
- Interest vector has a Gaussian cross-frequency 'shape' - A_k - and an azimuthal focus of A_{ear}.
- Initial connections have equal (maximum) strength. These decay to the attentional interest peak in an activity dependent manner. When a stimulus ceases, the connection strengths relax back to the maximum level.



$$ali = H\left(\sum_k H(x_k)\left[\left(\frac{\alpha_k}{\theta_\alpha}\right) - T_k\right]^+ - \theta_{ALI}\right) - ali \qquad T_k = 1 - (1 - A_{ear}A_k)L$$

where L is a leaky integrator and H is the unit step function. α_k is the gammatone instantaneous envelope for channel k and θ_k is a normalising factor. x_k is the oscillator activity at channel k.

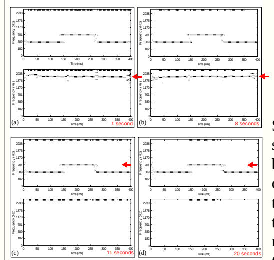
- Only the activity of oscillators whose connections fall under this attentional peak (and any synchronised oscillators) influence the ALI: the attended stream is synchronised with the activity of the ALI.
- It is proposed that L is subject to a 'reset' when the azimuthal focus, A_{ear}, changes. This results in an additional buildup period following a spatial movement of attention. Note that a reset does not occur for a movement of attention in frequency. This assumption is supported by psychophysical data collected by the authors.

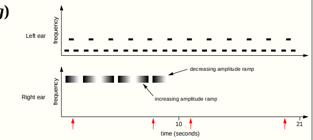


Results

Figures show a pseudo-spectrogram output from the model. Gray areas denote segments as created by the model. Each black pixel represents an oscillator in the active phase - an oscillator can only be active if it is part of a segment. The top row of black blocks represents ALI activity. Any active oscillators which are synchronised with the ALI are considered to be in the attentional foreground.

Two tone tone streaming (*auditory streaming*)

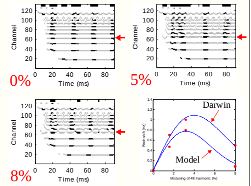




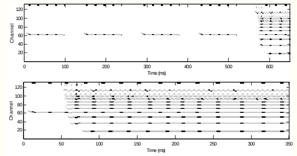
Snapshots of the model's response to an ABA sequence of tones (van Noorden, 1975) with noise burst distractors (Carlyon *et al.*, 2001) in the contralateral ear. Initially, attention is directed towards the noise bursts (figs. a and b). Streaming of the tones does not occur. After 10 seconds, attention moved to tones (fig. c). After the build up period (Anstis and Saida, 1985), streaming occurs (fig. d).

Mistuned harmonic and associated pitch shift (Darwin *et al.*, 1995)

Responses to complex tones. In each diagram, the complex tone has the fourth harmonic mistuned by a certain percentage. Below 8% (+50Hz), the mistuned harmonic remains grouped with the complex - all the oscillators are synchronised. Beyond 8%, the fourth harmonic segregates (represented by desynchronised activities). Associated with such segregation, listeners report a change in the pitch quality of the complex. This pitch *shift* reaches a maximum at approximately 4%. When the autocorrelation channels of the attended complex in the model are used to calculate the pitch, a quantitative match the pitch shift is seen.



Old-plus-new



The segregation of a harmonic by captor tones and asynchronous onset is accounted for by the incorporation of the *old-plus-new* heuristic: the auditory system's preference to '*interpret any part of a current group of acoustic components as a continuation of a sound that just occurred*' (Bregman, 1990). This segregation is indicated by the desynchronisation of the harmonic from the rest of the complex. Furthermore, this behaviour decays over time and eventually the harmonic re-fuses with the complex.

Summary

- A model of binaural auditory attention has been presented which is based on previous neural oscillator work by Wang and colleagues (Wang, 1996; Wang & Brown, 1999).
- Our model regards attention as a key factor in the stream formation process.
- We propose a 'reset' of attention when focus is moved spatially.
- Employs a unidimensional network which avoids a two dimensional time-frequency grid for which there is weak physiological justification.
- Sequential grouping is an emergent property of the model.
- Ability to incorporate other grouping cues to promote oscillator synchronisation.

References

Anstis, S. & Saida, S. (1985) Adaptation to auditory streaming of frequency-modulated tones. JEP:HPP 11 257-271.

Bregman, A. S. (1990) Auditory Scene Analysis. Cambridge MA: MIT Press.

Brown, G. J. & Cooke, M. (1994) Computational auditory scene analysis. Comput. Speech Lang. 8, pp. 297-336.

Carlyon, R. P., Cusack, R., Foxton, J. M. & Robertson, I. H. (2001) Effects of attention and unilateral neglect on auditory stream segregation. JEP:HPP 27(1) 115-127.

Darwin, C. J., Hukin, R. W. & Al-Khatib, B. Y. (1995) Grouping in pitch perception: Evidence for sequential constraints. J. Acoust. Soc. Am. 98(2)Pt1 880-885.

Moore, C.M. & Egeth, H. (1997). Perception without attention: evidence of grouping under conditions of inattention. JEP:HPP 23 339-352.

van Noorden, L.P.A.S. (1975). Temporal coherence in the perception of tone sequences. Doctoral thesis, Institute for Perceptual Research, Eindhoven, NL.

von der Malsburg, C. (1981). The correlation theory of brain function. Internal report 81-2, Max Planck Institute for Biophysical Chemistry, Gottingen, Germany.

Wang, D. L. (1996) Primitive auditory segregation based on oscillatory correlation. Cognitive Sci. 20 409-456.

Wang, D. L. & Brown, G. J. (1999) Separation of speech from interfering sounds based on oscillatory correlation. IEEE Trans. Neural Networks 10 684-697.