A Neural Oscillator Model of Auditory Selective Attention

Stuart N. Wrigley and Guy J. Brown

Speech and Hearing Research Group
Department of Computer Science
University of Sheffield, UK

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

Introduction

- **Sounds are generally heard in a mixture** - this needs to be separated to form a perceptual description of each sound source.

- **Auditory Scene Analysis** (ASA) (Bregman, 1990) performs this separation. Takes place in two conceptual stages:

1. Segmentation - acoustic mixture separated into its constituent ‘atomic’ units.

2. Grouping - units likely to have arisen from same source recombined.

⇒ One or more **Streams** - perceptual ‘objects’ produced by auditory grouping **over time**. Each stream describes a single sound source.

Binding Problem

How are features, which are represented in a distributed manner in the auditory nervous system, combined to form a meaningful whole?

Three broad categories of solution:

- Combinatorial — **Heirarchical organisation**
- Attentional — **Increase saliency of to-be-grouped features**
- Temporal correlation — **Temporal synchrony between to-be-grouped features**

Connectionist ASA

- Oscillatory correlation framework is a **temporal correlation** solution to the **binding problem**.
- Oscillators corresponding to grouped auditory elements are **synchronised**, and are **desynchronised** from oscillators encoding other groups.
- Supported by neurobiological findings:

Joliot *et al.*
40 Hz oscillations associated with auditory grouping

Llinás and Ribary
40 Hz activity synchronised over large distances

Barth and MacDonald
Auditory cortex oscillations modulated by thalamus

Attention in ASA

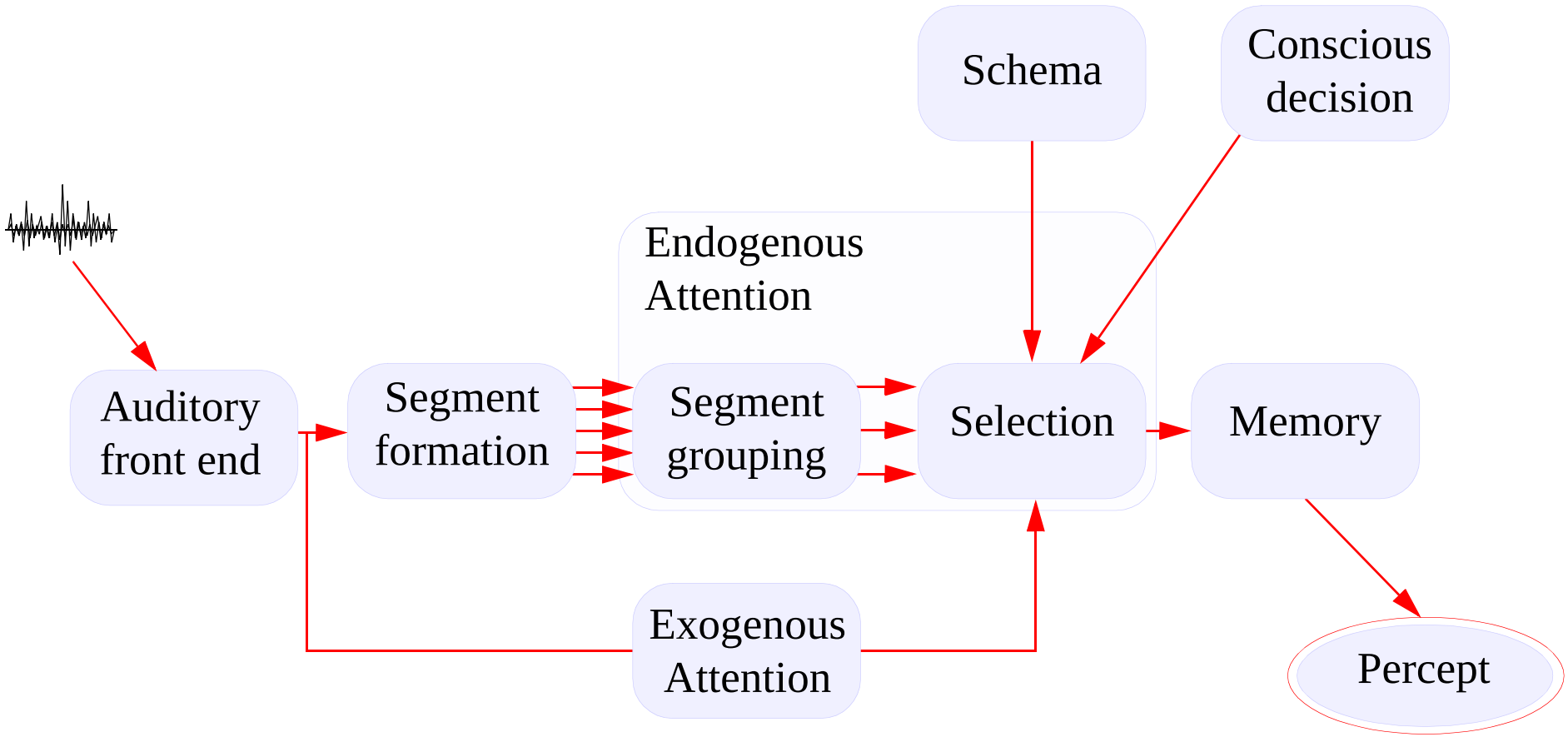
- Attentional research addresses **two core phenomena**:
  - **Capacity limits**
  - **Perceptual selectivity and its control**

- A number of **properties** to be modelled:
  - **Allocation**
  - **Shape**
  - **Two forms**
    - **Endogenous**
    - **Exogenous**

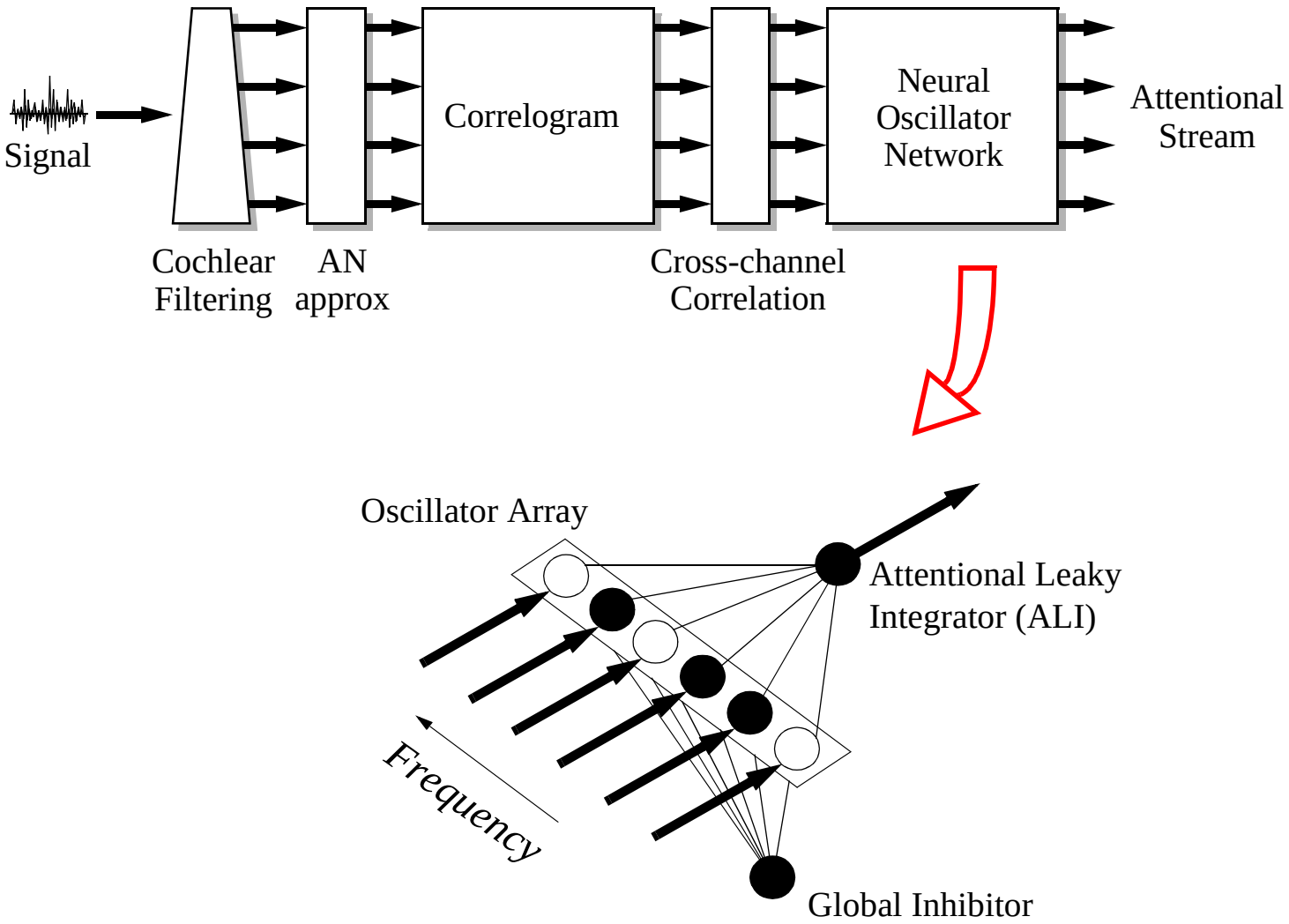
- Attention is **required for stream formation** and not just stream selection (Carlyon, 2001).

- Only attended streams are **encoded into memory** and perceived (e.g. Moore and Egeth, 1997).

Conceptual Model

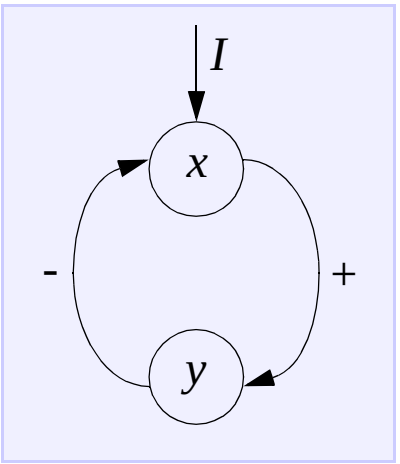


Oscillator-Based Implementation



Oscillators

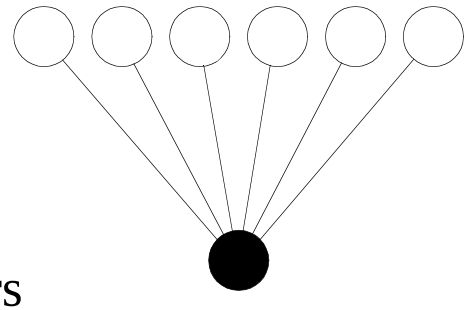
- *x* and *y* and each represent **mean activity** of a population of neurons.
- Population *y* acts as an inhibitor to the excitatory population *x* (See Terman and Wang, 1995).
- Oscillators corresponding to grouped auditory elements are **synchronised**, and are **desynchronised** from oscillators encoding other groups.



Attentional selection can be modelled by synchronising an attentional process to a particular group.

Oscillator Array

- Segment formation: excitatory and inhibitory connections promote **synchronisation** and **desynchronisation** of oscillator groups.
- Segments grouped on basis of harmonicity and stimulus ‘age’ (old-plus-new heuristic).
- Each oscillator in network feeds **excitatory input** to global inhibitor (a leaky integrator). Global inhibitor, in turn, feeds **inhibitory input** back to each oscillator.

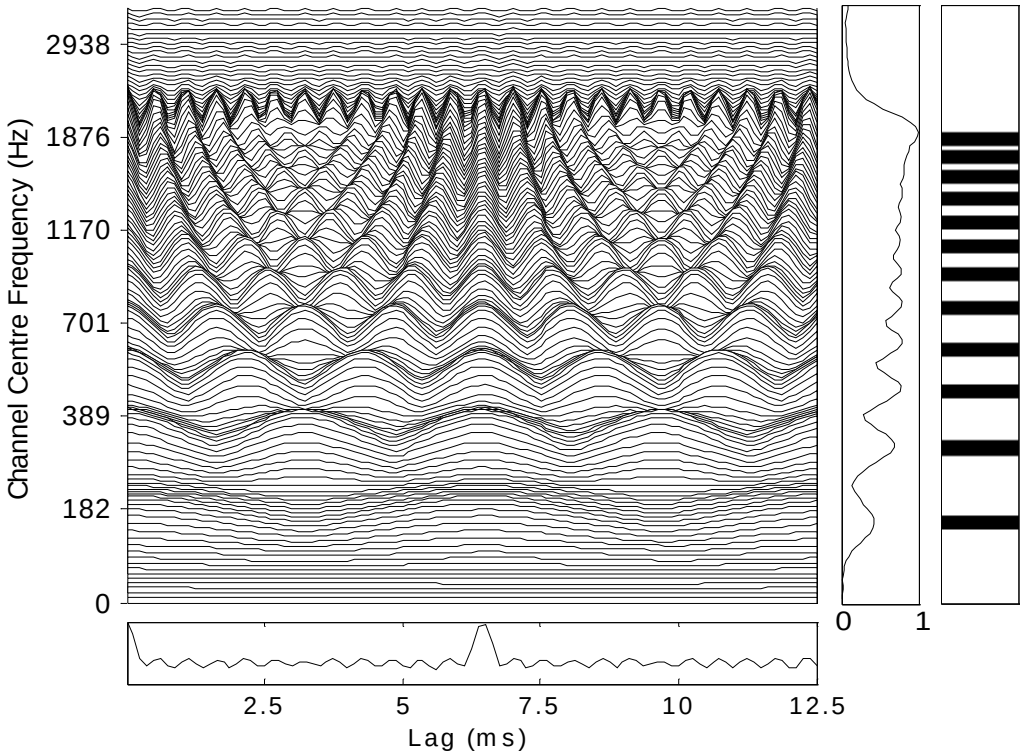


- When one group of synchronised oscillators is active, all others are **suppressed**; i.e. only one group can be active at any one time.

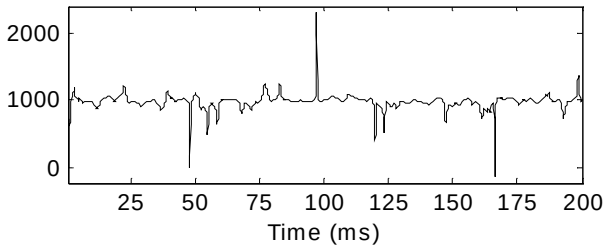
Segment formation

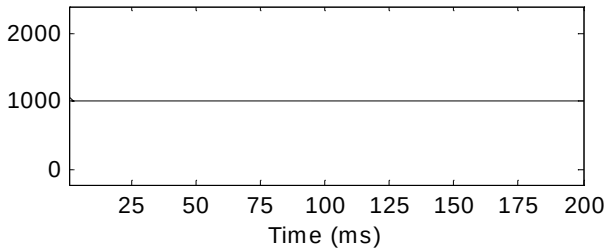
Two stage process:

- Common activity from correlogram extracted by cross-correlation of adjacent channels identifies channels stimulated by periodic energy.



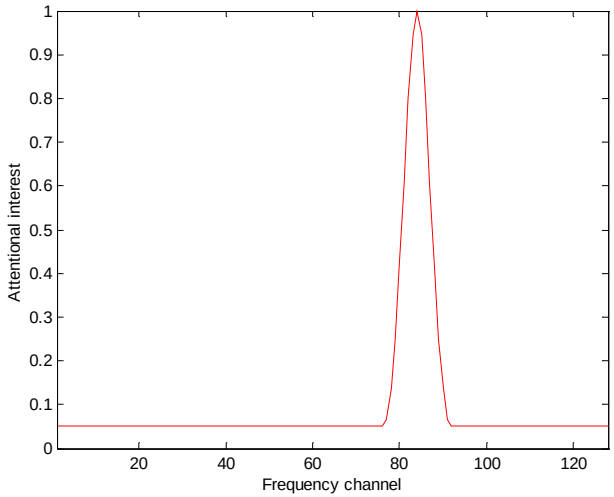
- Variance analysis of instantaneous frequency to identify channels stimulated by noise.





ALI: Stream Selection and Formation

- ALI connection strength determined by **'conscious' attentional interest** - a **Gaussian** in accordance with the gradient model of attention (e.g. Mondor and Zatorre, 1995).



- **Build-up** of streaming modelled by a slow leaky integrator to modulate connection strengths.

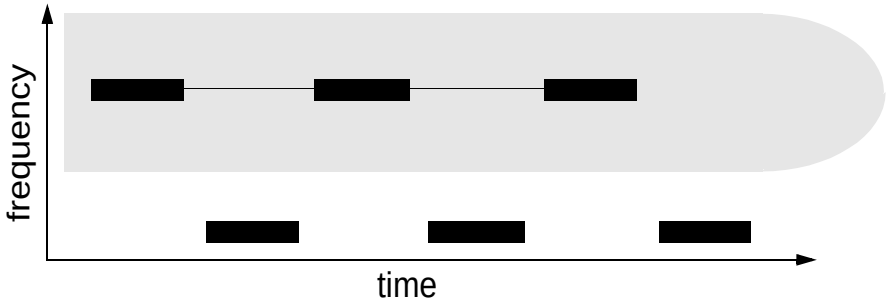
- Generally, only the activity of oscillators whose connections fall under this attentional peak influence the ALI.

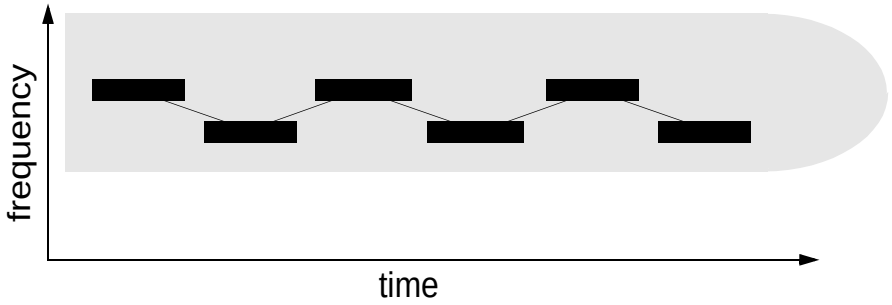
- ALI activity is **synchronised** with that of the attended stream.

ALI and Stream Formation

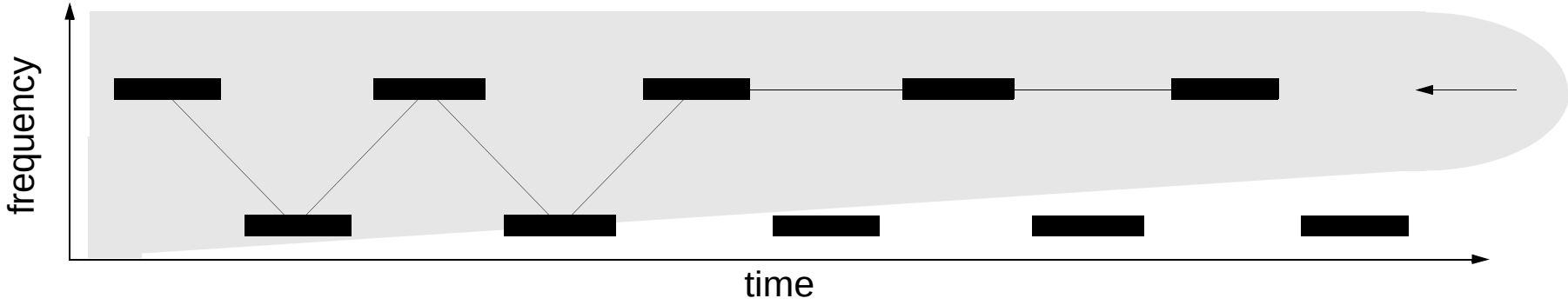
Example: Two Tone Streaming (van Noorden, 1975)

- Attentional allocation only



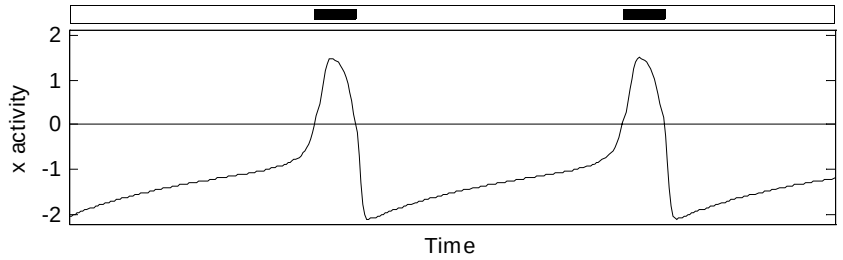


- Attentional allocation with build-up

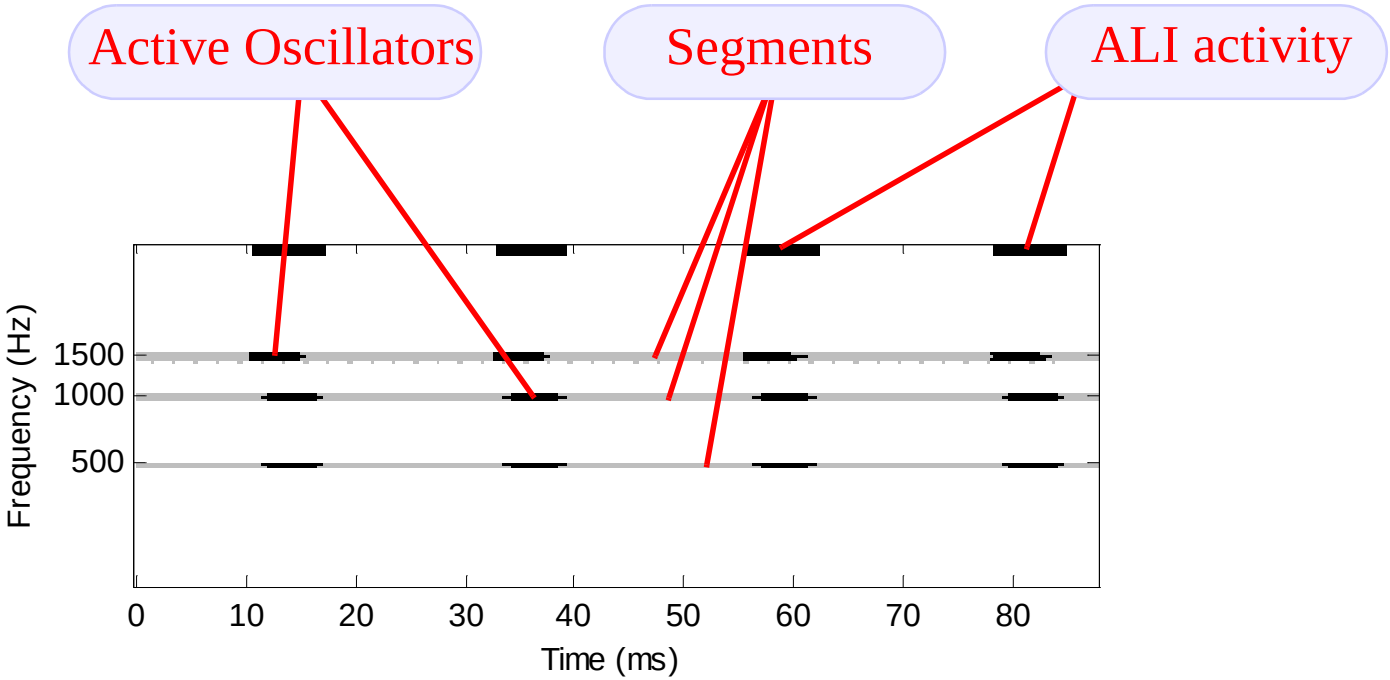


Output representation

- Oscillator activity

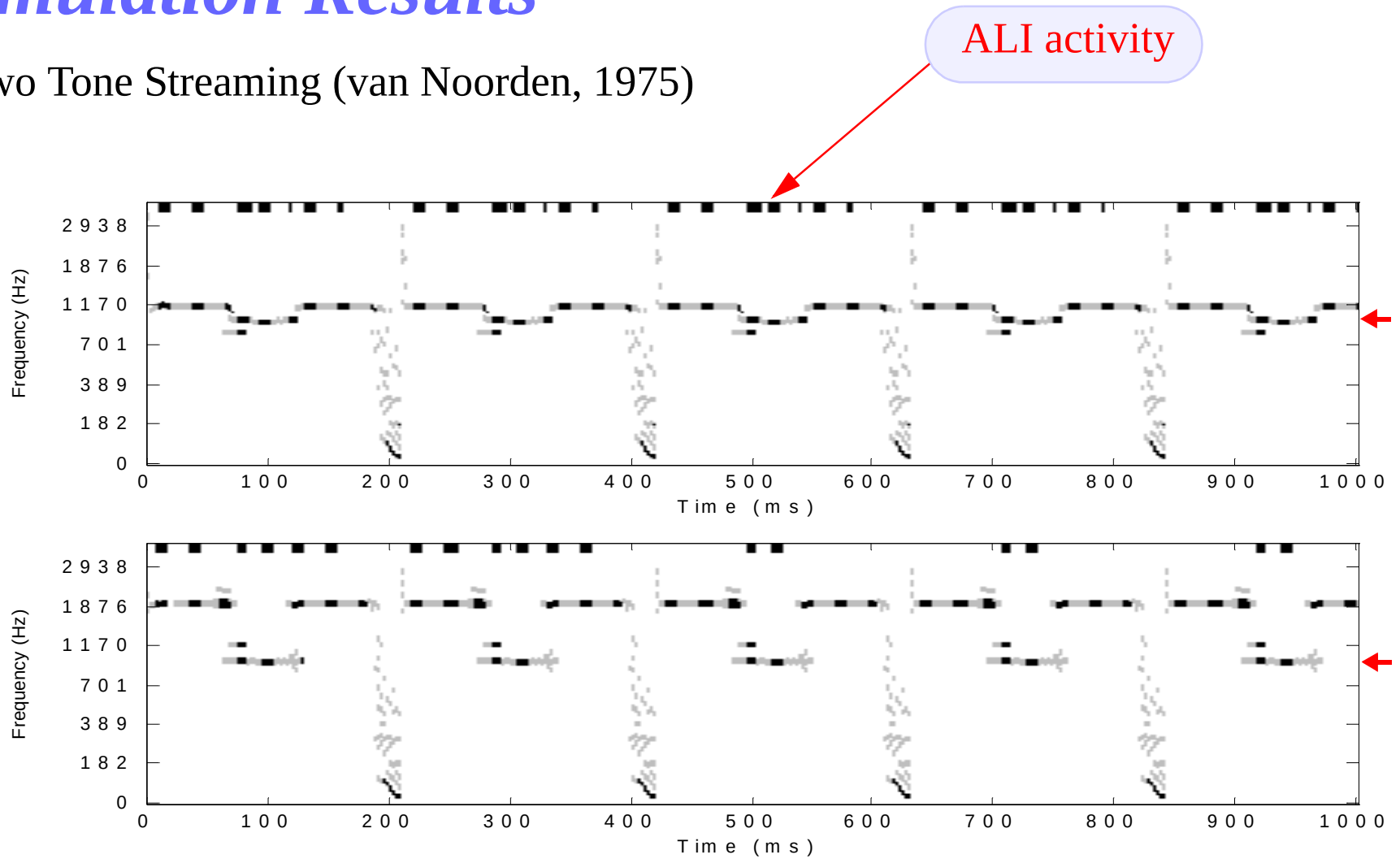


- Pseudospectrogram



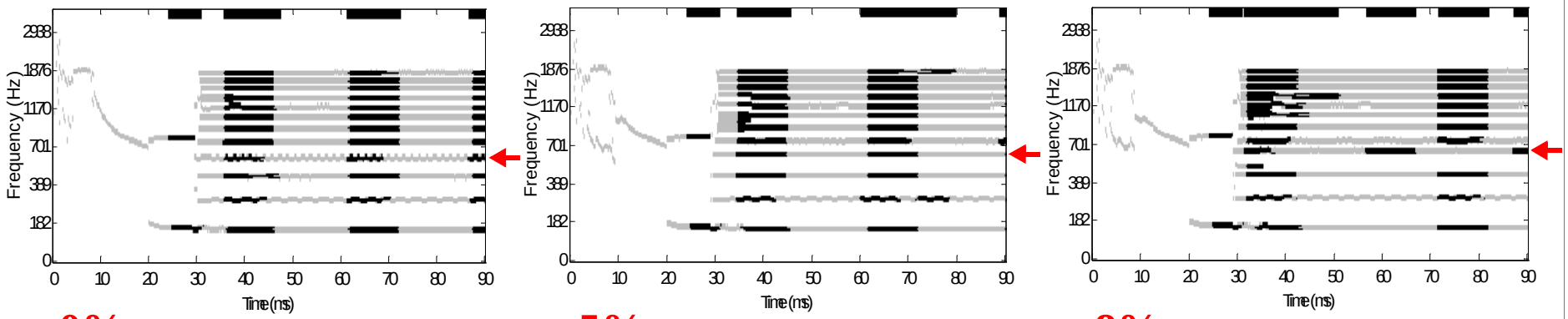
Simulation Results

- Two Tone Streaming (van Noorden, 1975)

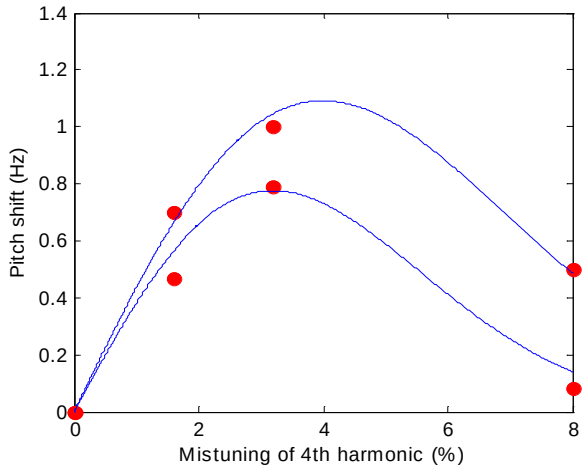


Simulation Results

- Mistuned harmonic

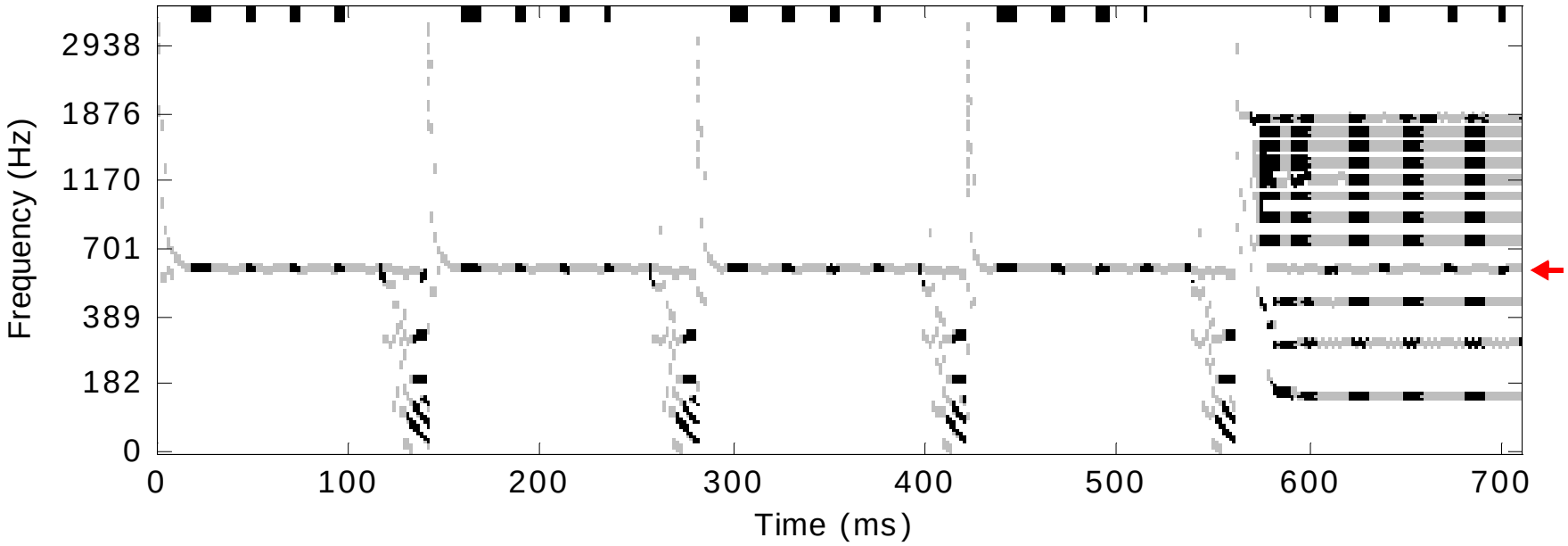


0% **5%** **8%**

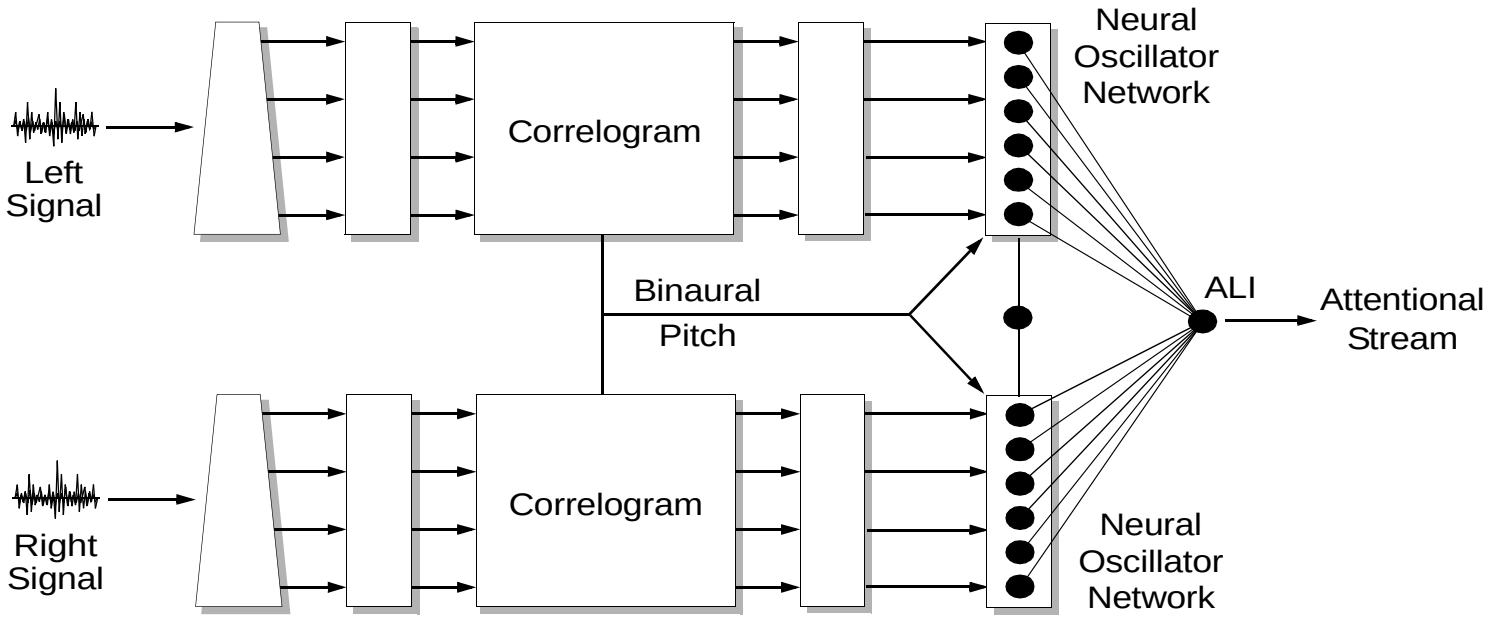


Simulation Results

- Tone capture (0% mistuning)



Binaural Implementation

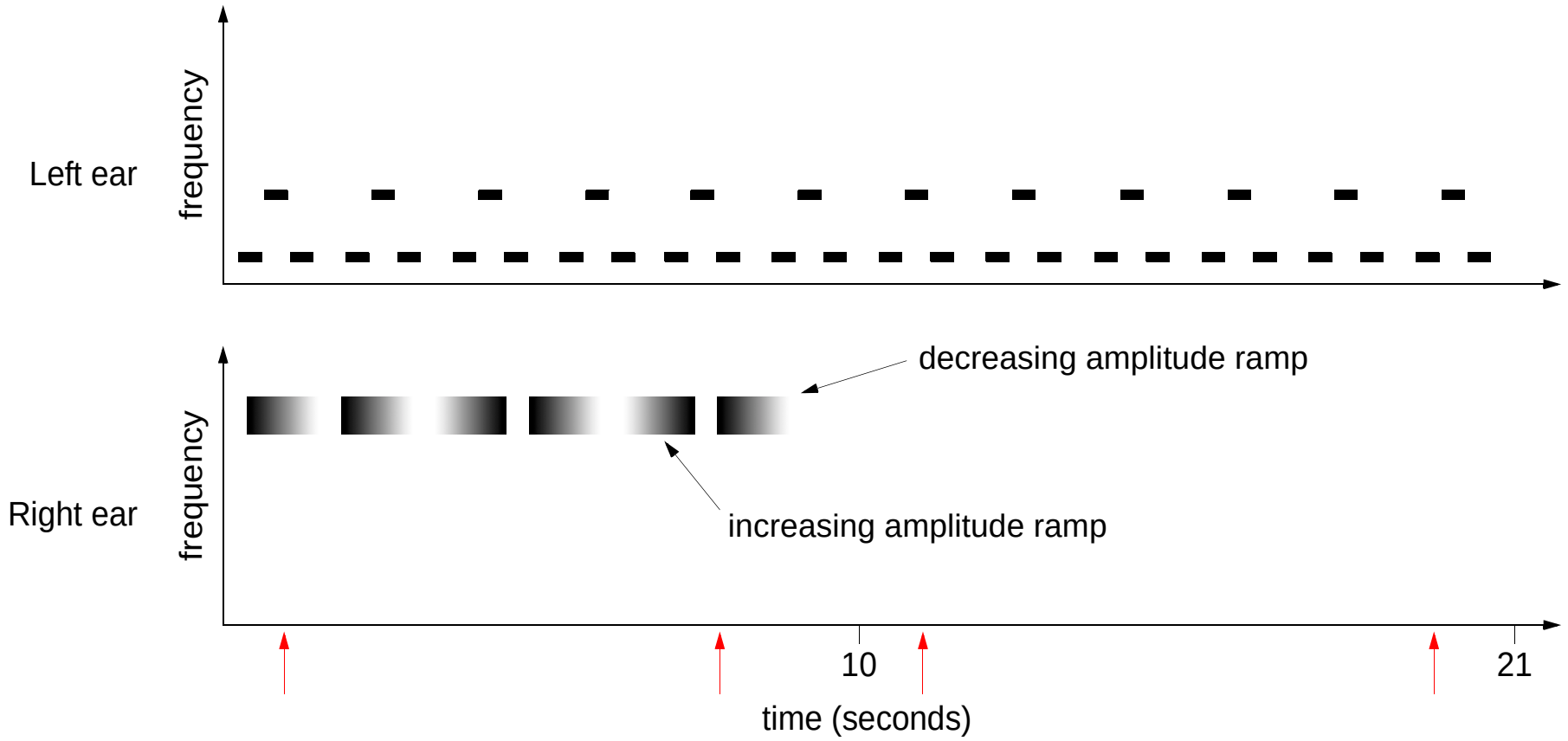


ALI same as monaural ALI but:

- Attentional interest vector A_k is modulated by a spatial interest weighting.
- Attentional build-up subject to a reset when change in spatial interest detected.

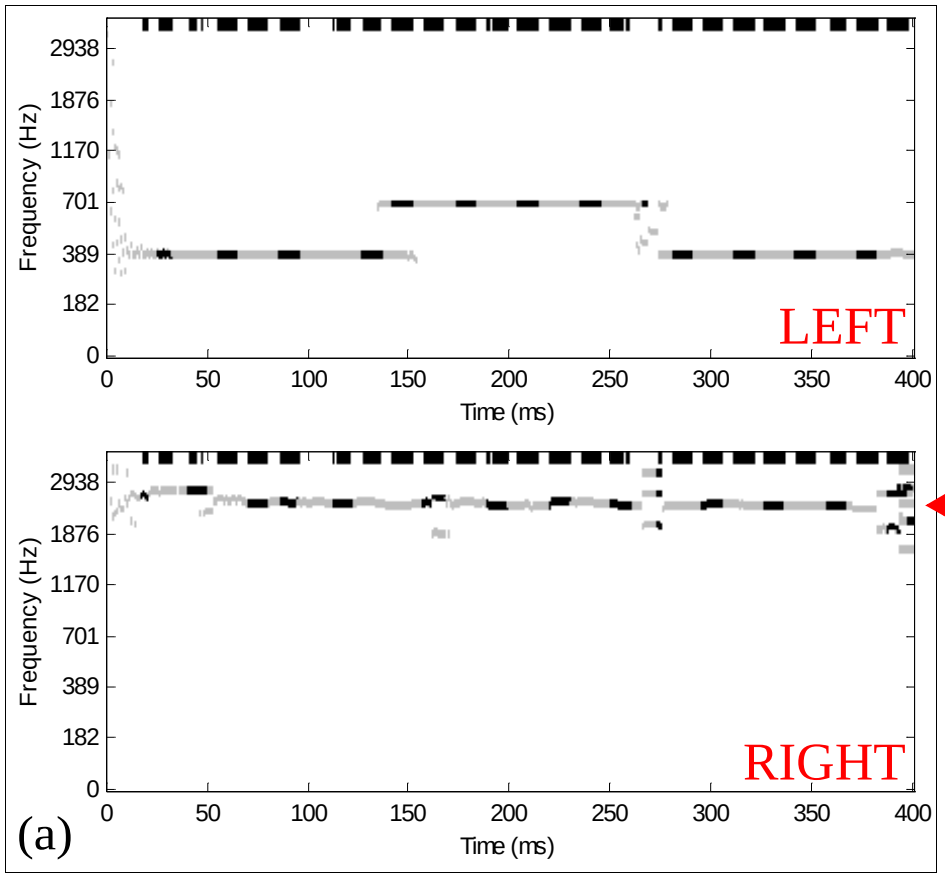
Binaural Simulation Results

- Two Tone Streaming with distractor (Carlyon et al., 2001)

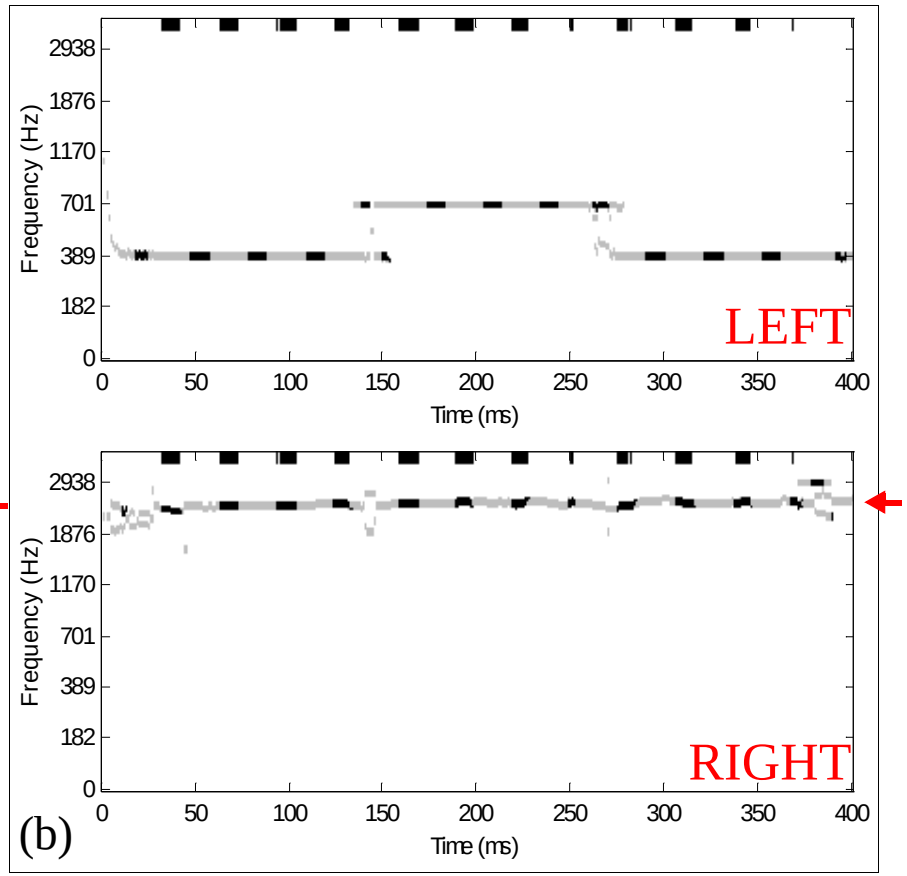


Binaural Simulation Results

- Two Tone Streaming with distractor (Carlyon et al., 2001)



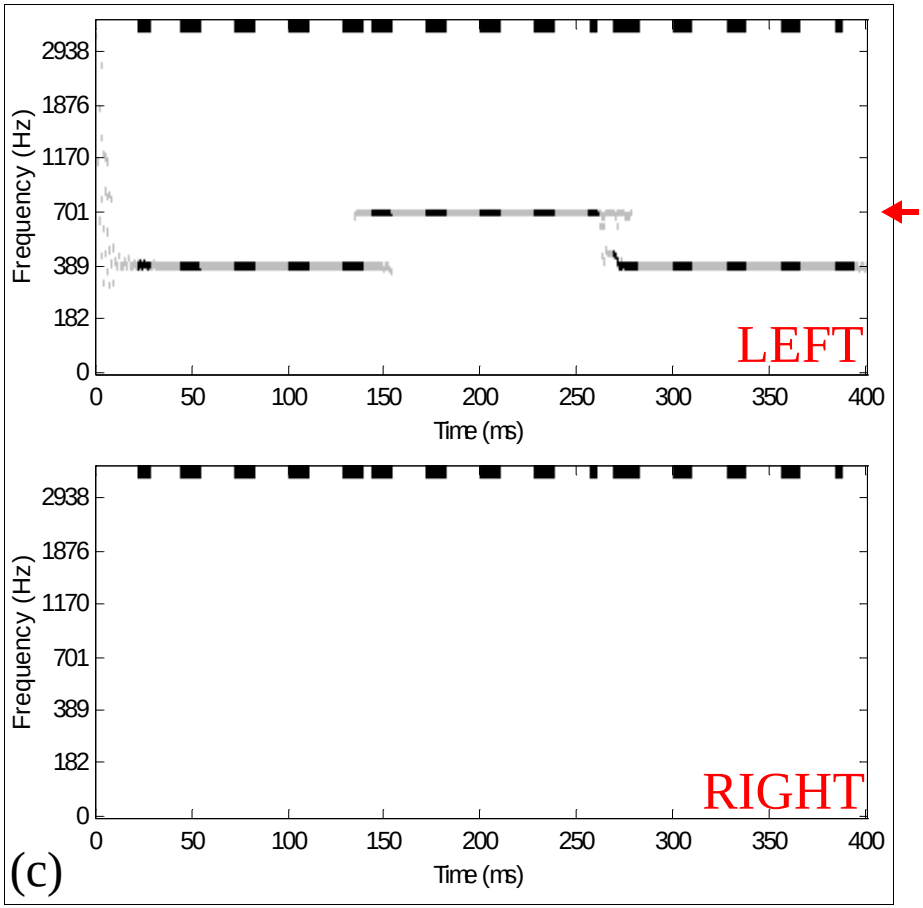
1 second



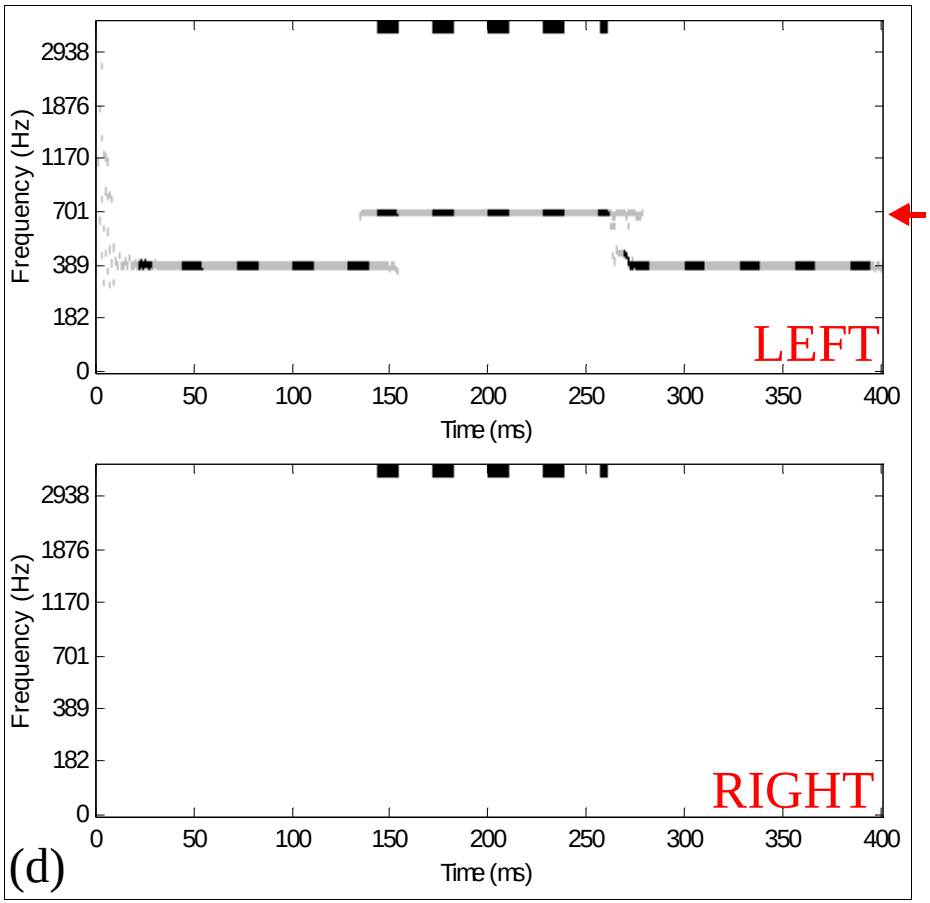
8 seconds

Binaural Simulation Results

- Two Tone Streaming with distractor (Carlyon et al., 2001)



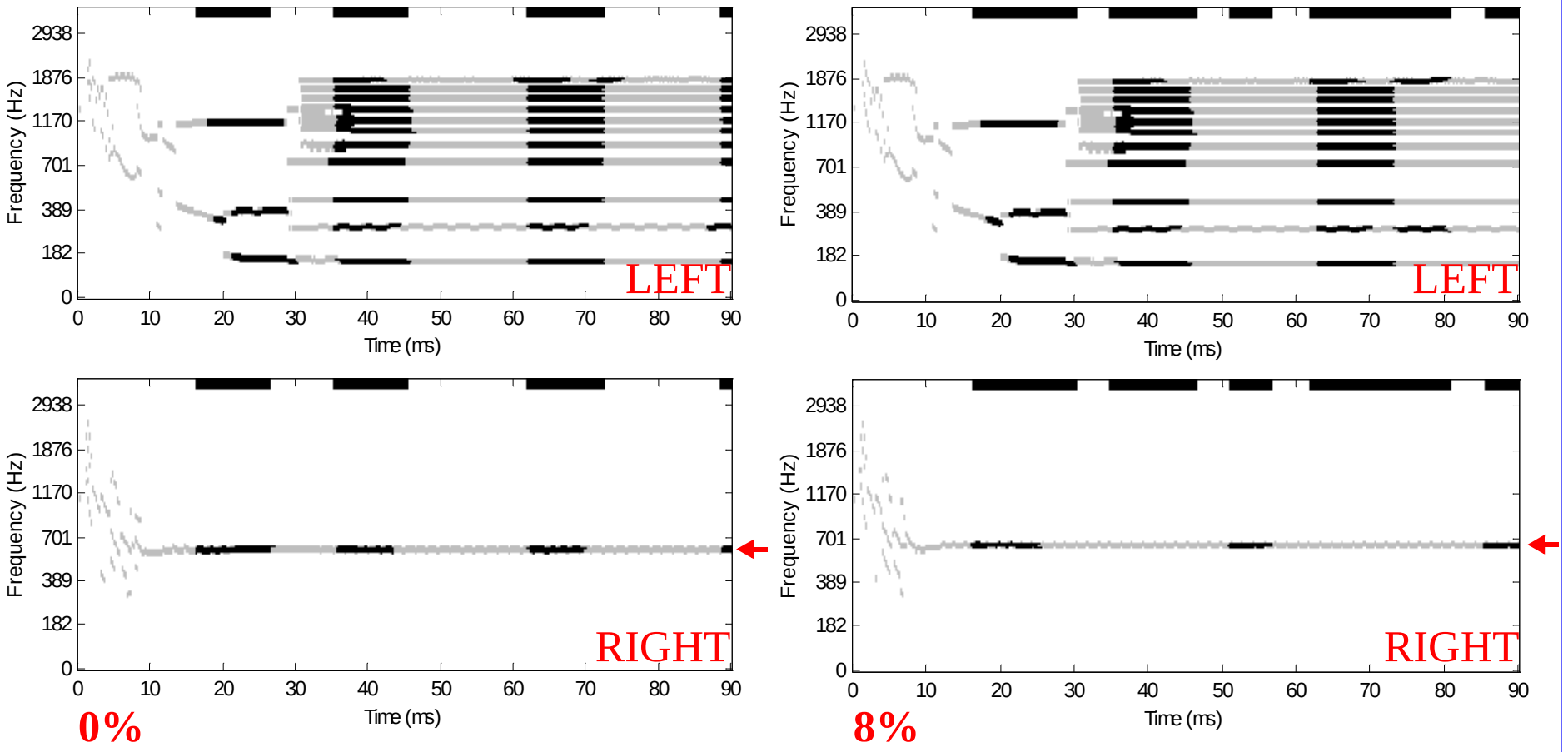
11 seconds



20 seconds

Binaural Simulation Results

- Complex tone and contralaterally presented mistuned harmonic (Darwin *et al.*, 1995)



Directions For Future Research

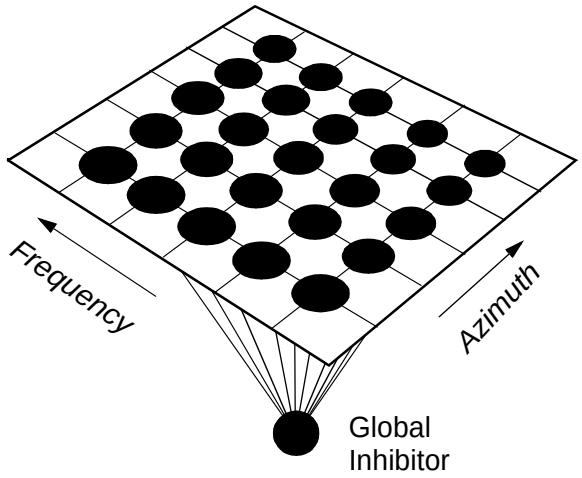
- Joint allocation: exclusive allocation enforced by the neural oscillator network. Can't account for duplex perception (e.g. Rand, 1974; Moore *et al.*, 1985).

  Brown and Wang (1996) work on double vowels: oscillators can be members of 2 groups.

- Divided attention: attention can be divided to multiple frequency regions simultaneously. Our model only incorporates selective attention.

- Binaural front end and neural oscillator network.

- Investigation into timecourse of binaural attentional allocation.



Summary

- A model of **attentionally modulated** auditory scene analysis has been presented.
- Uses an **oscillatory correlation** framework.
- Produces a **time-varying estimate of the attentional foreground**.
- Accounts for streaming of alternating tone sequences, build-up over time and the old-plus-new heuristic.
- Accounts for binaural grouping.
- Accounts for lack of streaming when attention directed elsewhere and non-transfer of streaming build-up when attention moved in space.