Auditory figure-ground segregation using a complex stochastic stimulus
Outline

- Introduction
- Stimulus
- fMRI study
- Psychophysics
- Temporal coherence model
Auditory scene analysis
Auditory figure-ground segregation

Listeners’ ability to extract a particular sound from a background of other simultaneous sounds

**Processes:**

i. grouping of simultaneous figure components from the spectral array,
ii. grouping of figure components over time,
iii. separation of grouped components from rest of the acoustic scene.

**Mechanisms:**

- Segregation is mediated by basic response properties of auditory cells: frequency selectivity, forward suppression and adaptation, resulting in the activation of distinct neuronal populations

**Stimuli:**

- Studied using relatively simple signals, which lack the rich spectrotemporal complexity of natural signals, e.g. streaming signals
Streaming

ABA_…ABA_ or AB…AB

\[ \Delta F: 1 \text{ st} \quad \text{Integrated percept} \]

\[ \Delta F: 6 \text{ st} \quad \text{Ambiguous percept} \]

\[ \Delta F: 9 \text{ st} \quad \text{Segregated percept} \]
Stochastic Figure-Ground stimulus

A  No figure

Time (ms)

Hz

9446
3027
867
140
SFG: Figure present

B Figure with ‘coherence’ = 4 and ‘duration’ = 7
SFG: Stimulus design

**Stimulus:**
Sequence of random chords consisting of pure tone components

Duration of each chord: 50 ms
Inter-chord interval: 0 ms
Total stimulus duration: 2000 ms (40 consecutive chords)

**Chords:**
No. of pure tone components: 5-15
Component frequency range: 179 – 7246 Hz
Cosine ramp: 10 ms for onset and offset

**Coherence:**
Number of different repeating frequencies: 1,2,4,6,8

**Duration:**
Number of chords over which frequencies repeat: 2-7
SFG: Figure

Features of SFG:

• Figure and background signals do not differ in low-level acoustic attributes

• No spectral ‘protective’ region between figure and background

• Figure and background signals are indistinguishable at each point in time

• Figure can only be extracted by integrating over time and frequency

• Enables parametric variation of figure salience
Listeners are remarkably sensitive to the appearance of figures.

Sensitive to parametric variations of coherence and duration.
Aim: Identify brain areas whose activity varies with parametric variations in coherence and duration of the figure.

Stimulus: 
  i. Fixed coherence: 4, varying duration: 2-7 chords
  ii. Fixed duration: 4, varying coherence: 1,2,4,6,8

  = 9 stimulus conditions (x 40 repetitions)

Paradigm: 
  i. Passive listening
  ii. Active figure-detection
We used the SFG stimulus in a passive fMRI study to identify brain areas whose activity varies parametrically with coherence and duration of the figure.

**Task:** Detect decoy stimuli (noise bursts; 10% of stimuli)

**Results:** We found activity in the **intraparietal sulcus** (cf Cusack 2005) and **superior temporal sulcus** to be involved in automatic, stimulus-driven, bottom-up segregation, with no role for the primary auditory cortex.

*Teki et al., J Neurosci (2011)*
Aims

• To characterize the brain mechanisms that underlie complex figure-ground segregation through systematic manipulations of the SFG stimulus
Expt. 1: ‘Baseline’ (50 ms)

Stimulus consisted of a sequence of 40 x 50ms chords (2 s long)

Coherence: [1 2 4 6 8]      Duration: [2:7]
Expt. 1: Results

(n=9)
Expt. 2: ‘Baseline’ (25 ms)

Stimulus consisted of a sequence of 40 x 25ms chords (1 s long)

Coherence: [2 4 6 8]  Duration: [2:10]
Expt. 2: Results
(n=8)
Expt. 1 vs. 2

ANOVA

- Coherence and duration as within-subject factors
- Chord length (50 ms vs. 25 ms) as between-subject factor

Results

**Significant** effect of coherence: \( F(3, 45) = 77, p < 0.001 \)

**Significant** effect of duration: \( F(5, 75) = 41, p < 0.001 \)

No significant effect of chord length: \( F(1,15) = 2, p = 0.174 \)
Expt. 3: ‘SFG/Noise’

Stimulus: SFG with 40 x 50ms chords alternating with 50ms of white noise (4 s)

Expt. 3: Results
(n=10)
Expt. 1 vs. 3

ANOVA

- Coherence and duration as within-subject factors
- Condition (Baseline vs. SFG/Noise) as between-subject factor

Results

**Significant** effect of coherence: $F(3, 51) = 23, p < 0.001$

**Significant** effect of duration: $F(4, 68) = 29, p < 0.001$

No significant effect of condition: $F(1,17) = 0.004, p = 0.953$
Stimulus: Figures were ramped (successive figure components were not repeating but increasing in frequency in steps of $2I$ or $5I$, where $I = 1/24$ of an octave is the resolution of our frequency pool; ramps within critical band)

Coherence: [4 6 8]  
Duration: [5 7 9]  
Ramp step: [2/5]
Results: Ramps 2

(n=10)

![Graph showing the relationship between D prime and duration (in chords) for different Coh levels (Coh=4, Coh=6, Coh=8).]
Results: Ramps 5
(n=10)
Expt. 1 vs. 4a vs. 4b

ANOVA

- Coherence (4, 6, 8) and duration (5, 7) as within-subject factors
- Condition (Baseline vs. ramp of 2 vs. ramp of 5) as between-subject factors.

Results

**Significant** effect of coherence: \( F(2, 50) = 25, p < 0.001 \)

**Significant** effect of duration: \( F(1, 25) = 110, p < 0.001 \)

**Significant** effect of condition: \( F(2,25) = 19, p < 0.001 \)
Expt. 4a vs. 4b

ANOVA

- Coherence (4, 6, 8) and duration (5, 7) as within-subject factors
- Condition (Ramp of 2 vs. Ramp of 5) as between-subject factors.

Results

**Significant** effect of coherence:  \( F(2, 36) = 70, p < 0.001 \)

**Significant** effect of duration:  \( F(2, 36) = 198, p < 0.001 \)

**Significant** effect of condition:  \( F(1,18) = 21, p < 0.001 \)
Expt. 5: ‘Isolated’

Stimulus consisted only of the chords comprising the figure, and the preceding as well as succeeding chords were removed.

Coherence: [2 4 6 8]  
Duration: [3:9]
Expt. 5: ‘Isolated’
(n=10)
Expt. 1 vs. 5

ANOVA
- Coherence and duration as within-subject factors
- Condition (Baseline vs. Isolated) as between-subject factor

Results

Significant effect of coherence: $F(3, 48) = 85, p < 0.001$

Significant effect of duration: $F(4, 64) = 69, p < 0.001$

No significant effect of condition: $F(1,16) = 0.033, p = 0.859$
Summary

SFG figure-detection performance is:

- Depends on no. of repeating chords, not duration of figure (Expt. 1 & 2)
- Invariant to disruption by white noise (Expt. 1 & 3)
- Sensitive to shape of figure (continuous vs. ramped) (Expt. 1 & 4)
- Sensitive to size of ramps (2 vs. 5) (Expt. 4a & 4b)
- Invariant to the presence of preceding background (Expt. 1 & 5)
Listeners can segregate figure from background in the SFG stimulus very well.

Results suggest that adaptation is not critical for figure-ground segregation.

We found the behaviour to be congruent with the temporal coherence model of auditory scene analysis which suggests that auditory segregation is based on the computation of cross-channel coherence.

(Elhilali et al., 2009; Shamma et al., 2011)
Temporal coherence model
The temporal coherence model incorporates three different stages:

1. cochlear processing
2. processing of the signal at the cortical level
3. cross-correlation analysis of multi-scale, multi-rate cortical representations

The cortical model simulates A1 cells that are tuned to a range of spectral and temporal resolutions. Here, we modelled A1 cells tuned to a rate of 20 Hz with a bandwidth of 24 channels per octave.

Analysis:

The model was run for 1000 iterations for each stimulus condition for the figure and ground stimuli separately.

The maximum value of cross-correlation was computed for each stimulus and averaged across iterations to produce the model output.

The average value for ground stimuli was subtracted from the average value for the figure stimuli to obtain the model response.
Isolated: figure with coh=6, dur=6
Isolated: ground with coh=6, dur=6
Output: cross-correlation matrix
Ground: cross-correlation matrix
Expt. 5: Model

sv=24, rv=20
Expt. 1: Model

sv=24, rv=20

[Graphs showing D prime and Cross Correlation as functions of duration (in chords) for different Coh levels.]
Expt. 2: Model

sv=24, rv=40
Acknowledgments

Deborah, Sukhbinder and Tim  
Newcastle Auditory Group

Maria, Aiysha and Nicolas  
UCL Ear Institute

Shihab Shamma  
University of Maryland, College Park
Aim: Identify brain areas whose activity varies with parametric variations in coherence and duration of the figure.

Stimulus:
1. Fixed coherence: 4, varying duration: 2-7 chords
2. Fixed duration: 4, varying coherence: 1,2,4,6,8

Task: Detect decoy stimuli (noise bursts; 10% of stimuli)
- Subjects were not actively detecting figures
Behaviour in scanner

![Graph showing the relationship between hit rate and parametric modulation with different coherence and duration levels.]

Key:
- Coherence (fMRI)
- Duration (fMRI)
- Coherence (Psy)
- Duration (Psy)
- False positive rate

Parametric modulation:
- Coherence: 1, 2, 4, 6, 8
- Duration: 2, 3, 4, 5, 6

Hit rate range: 0.3 to 1.0
Effects of Duration

A
- Left IPS
  - y = -46
- Right IPS
  - y = -28

B
- Left STS
  - y = -34
- Right STS
  - y = -13
Effects of Coherence

A

Left IPS

Right IPS

\( y = -73 \)

\( y = -82 \)

B

Left STS

Right STS

\( y = -16 \)

\( y = -4 \)
What about the auditory cortex?

- No activation in Primary Auditory Cortex (PAC) for either contrast
- Confirmed using volume of interest analysis based on PAC maps (Morosan et al., 01)
- Consistent with one previous fMRI study (Cusack, 2005)

Reasons...
- More complex and naturalistic stimulus
- Naïve subjects and short figures
- PAC recruited during active figure-ground segregation (i.e., in behavioural context) with possibly top-down modulation by IPS?

Role of STS
- STS activity modulated by changing duration and coherence of the figure

- Implicated in:
  - analysis of spectral shape (Warren et al., 2005)
  - dynamic changes in spectrum (Overath et al., 2008)
  - detection of increasing changes in spectrotemporal coherence within textures (Overath et al., 2010)
IPS and Perceptual Organization

Role of IPS consistent with Cusack (2005):
- Implicated IPS in perception of two streams vs. one stream, based on the same physical streaming signal that evoked a bistable percept.
- IPS activity likely reflects top-down application of attention (shift between streams)
- Found no activation in primary auditory cortex

IPS is involved in structuring sensory input and perceptual organization:
- Encoding visual object representations
- Binding of sensory features within and across different modalities
- Control and shift of auditory attention

What does the IPS activity reflect?
➢ automatic, bottom-up segregation of auditory object from stochastic background
Summary

SFG stimulus
• More representative of the natural complexity of acoustic scenes
• Figure can only be extracted by integrating over frequency-time space
• Shorter build up time (~300ms; compared to ~2s for streaming stimuli)
• Enables parametric approach to study auditory figure-ground segregation

Substrates
• IPS and STS: pre-attentive, stimulus-driven, bottom-up segregation
• No role of primary auditory cortex in such bottom-up segregation

Questions...
• Is IPS involved in active figure-ground segregation? And PAC?
• Functional connectivity between IPS and the auditory system?

For complete details, see: Teki, Chait et al., J Neurosci (2011)