

Sundeep Teki¹ Maria Chait³
Deborah Williams² Aiysha Siddiq³
Nicolas Barascud³ Sukhbinder Kumar^{1,2}
Shihab Shamma⁴ Tim Griffiths^{1,2}



Auditory figure-ground segregation using a complex stochastic stimulus

- ¹ Wellcome Trust Centre for Neuroimaging, University College London, UK
- ² Newcastle Auditory Group, Newcastle University Medical School, UK
- ³ UCL Ear Institute, University College London, UK
- ⁴ Institute for Systems Research, University of Maryland, College Park, USA

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:

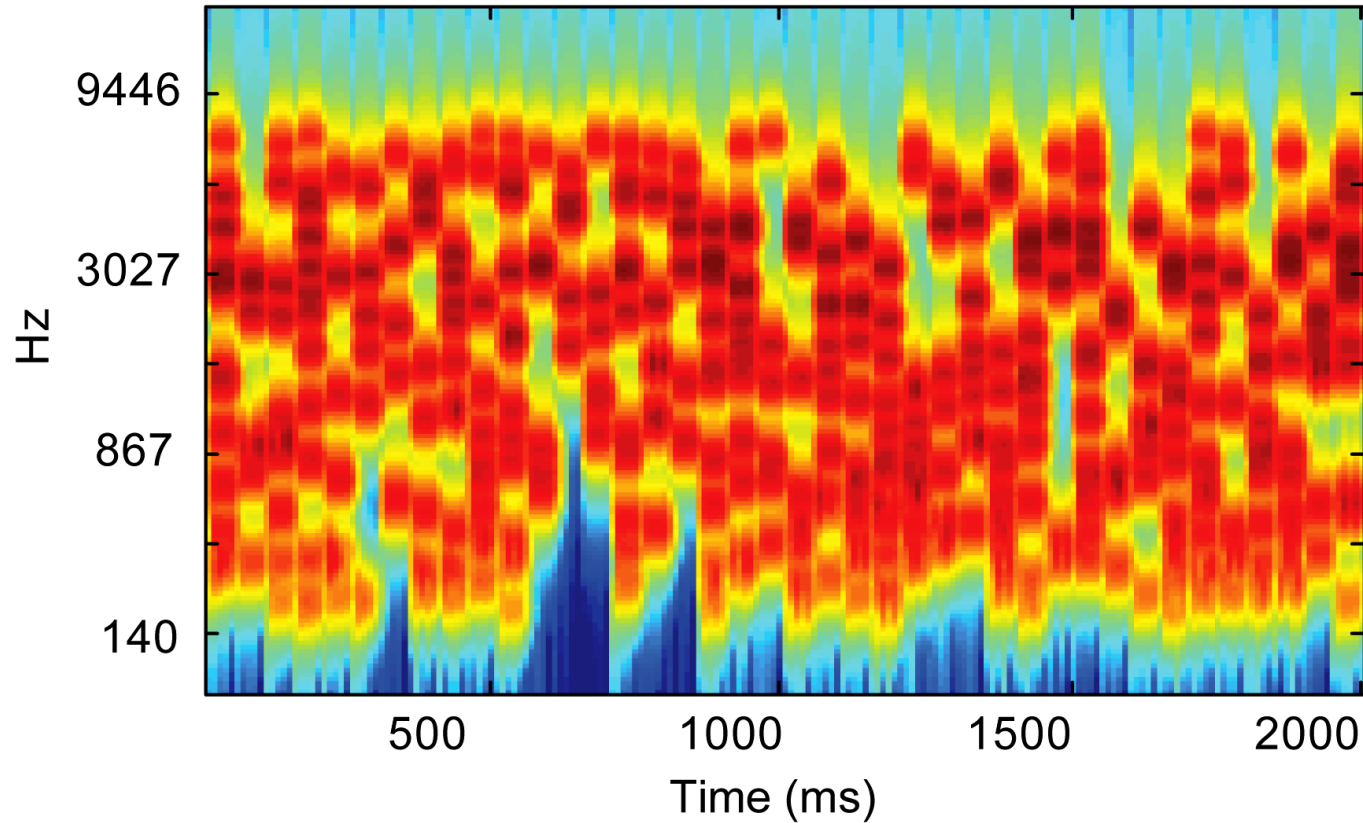
- Segregation is mediated by basic response properties of auditory cells: frequency selectivity, forward suppression and adaptation, resulting in the activation of distinct neuronal populations

Stimuli:

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

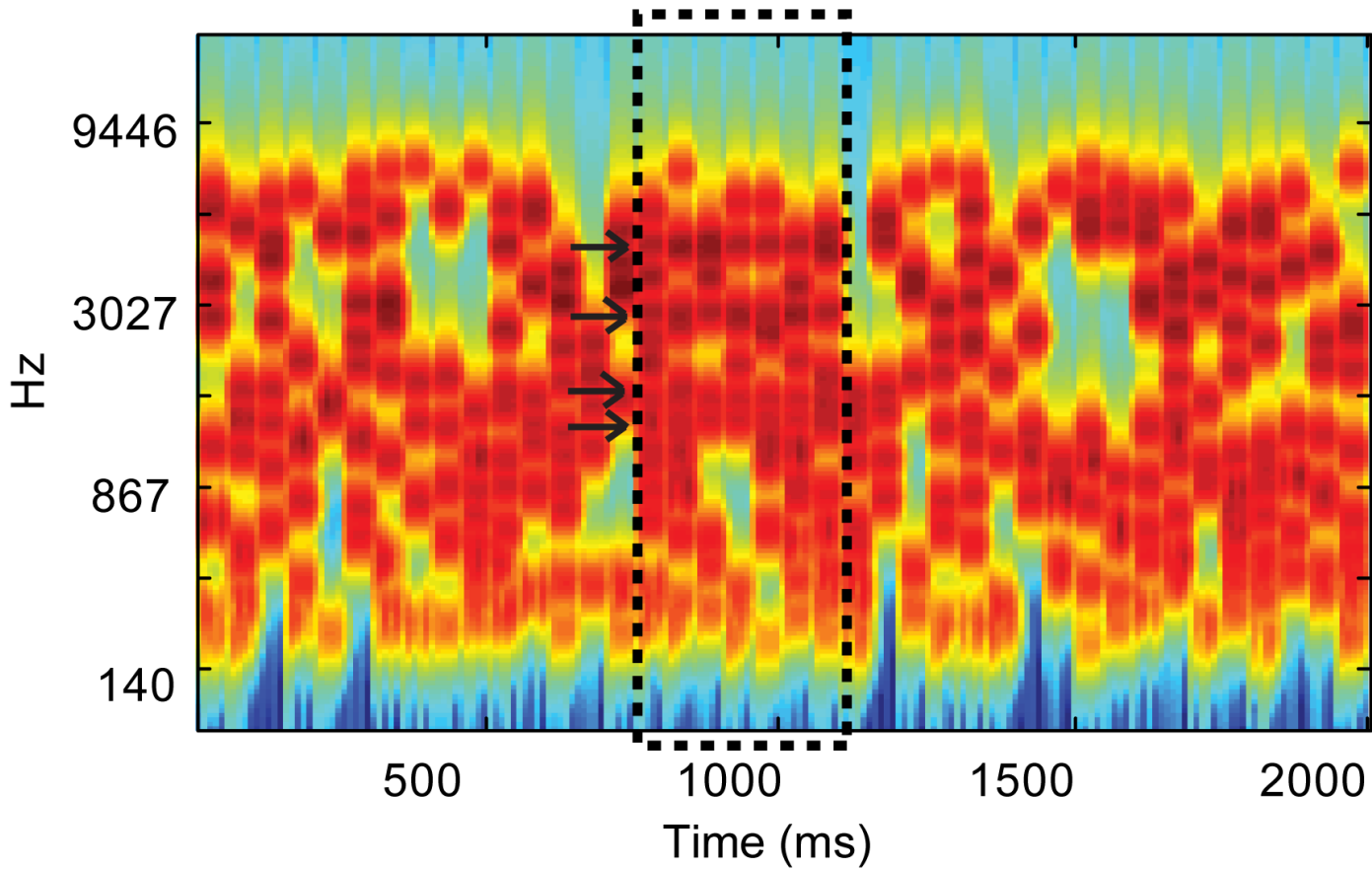
Stochastic Figure-Ground stimulus

A No figure



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
--	------------

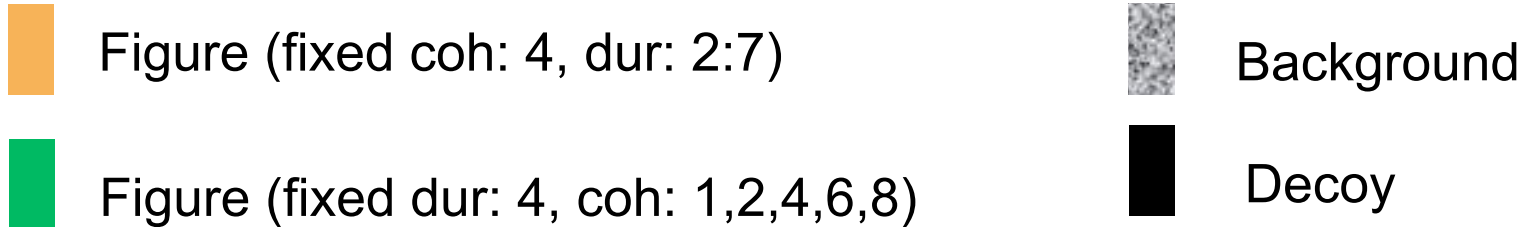
SFG: Figure

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

fMRI experiment

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 (IPS) and superior temporal sulcus (STS) to be involved in automatic, stimulus-driven, bottom-up segregation, with no role for the primary auditory cortex.

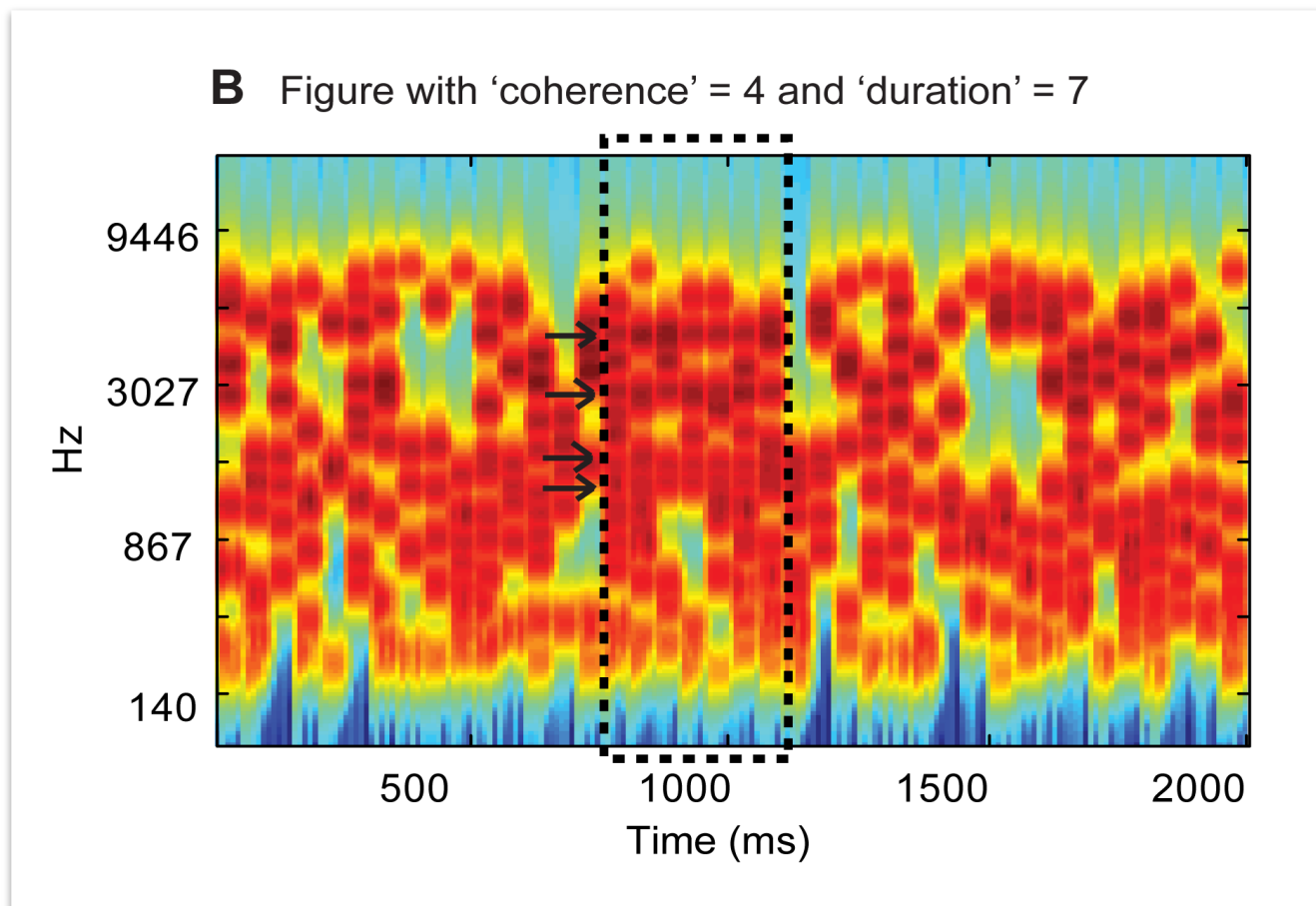
Teki et al., J Neurosci (2011)

Aim of present study

To characterize the brain mechanisms that underlie figure-detection performance, and elucidate the perceptual organization of complex acoustic scenes, we used psychophysics to investigate the degree to which performance is affected by systematic manipulations of the spectro-temporally complex SFG stimulus.

Expt. 1: 'Baseline' (50 ms)

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

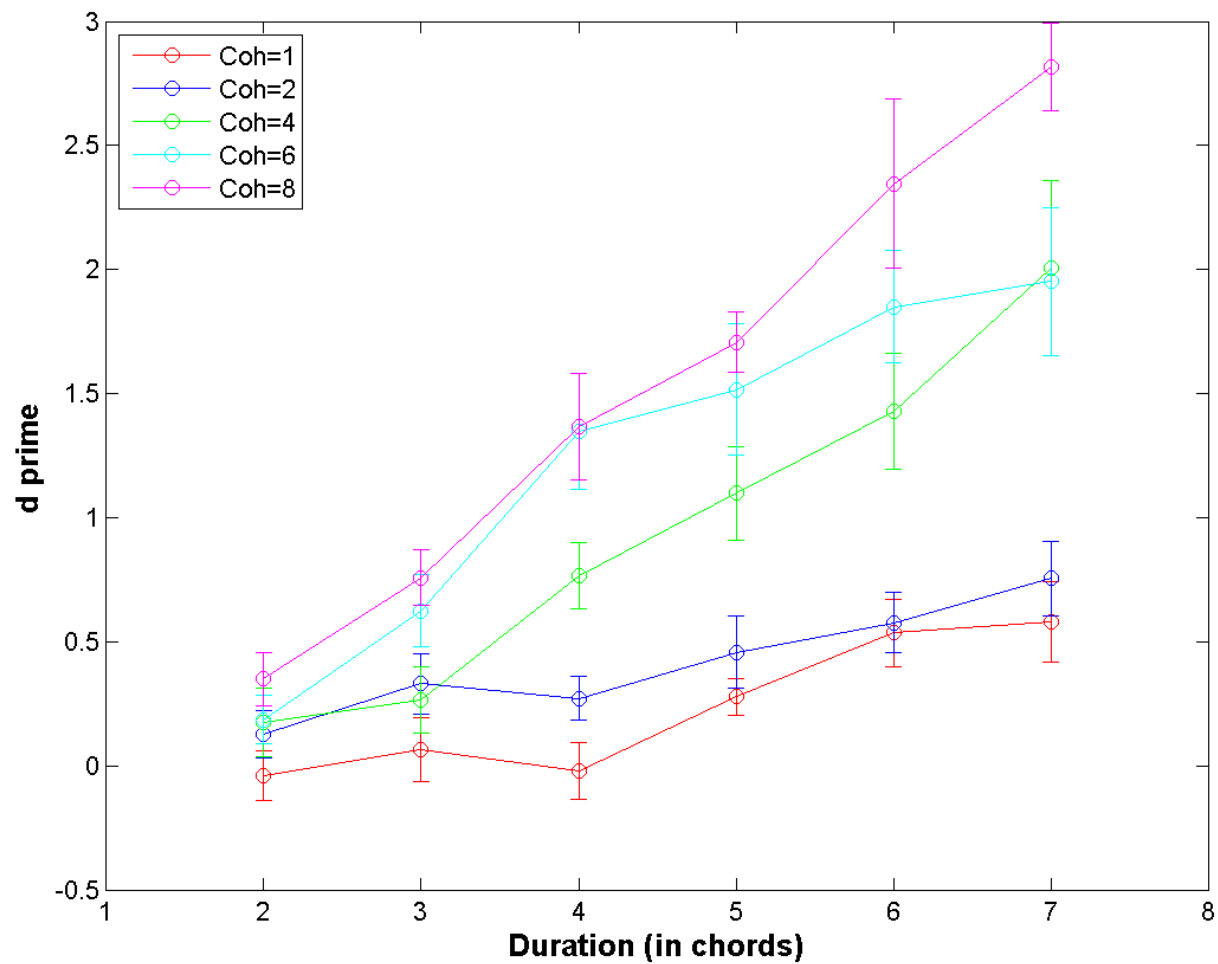


Coherence: [1 2 4 6 8]

Duration: [2:7]

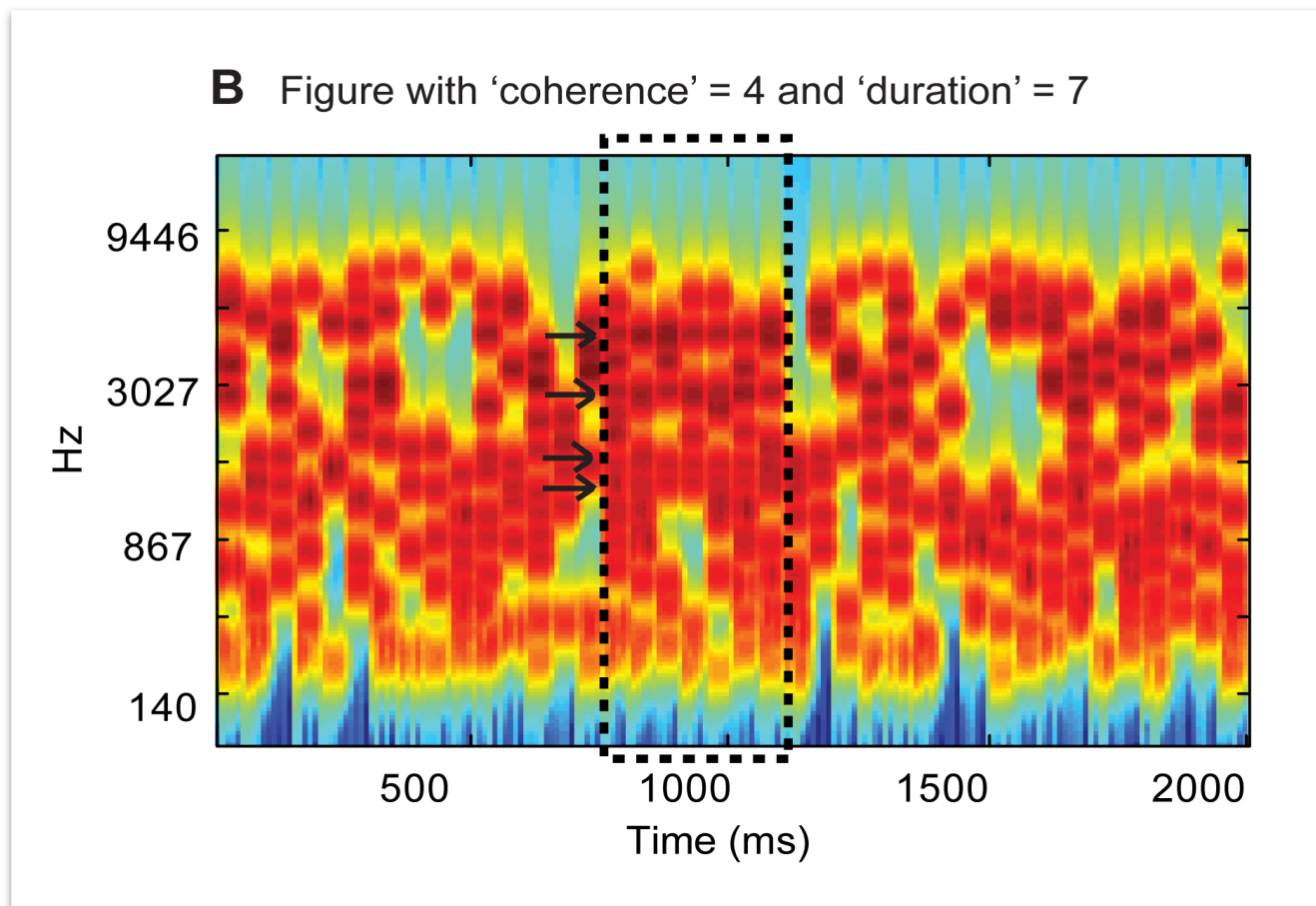
Expt. 1: Results

(n=9)



Expt. 2: 'Baseline' (25 ms)

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

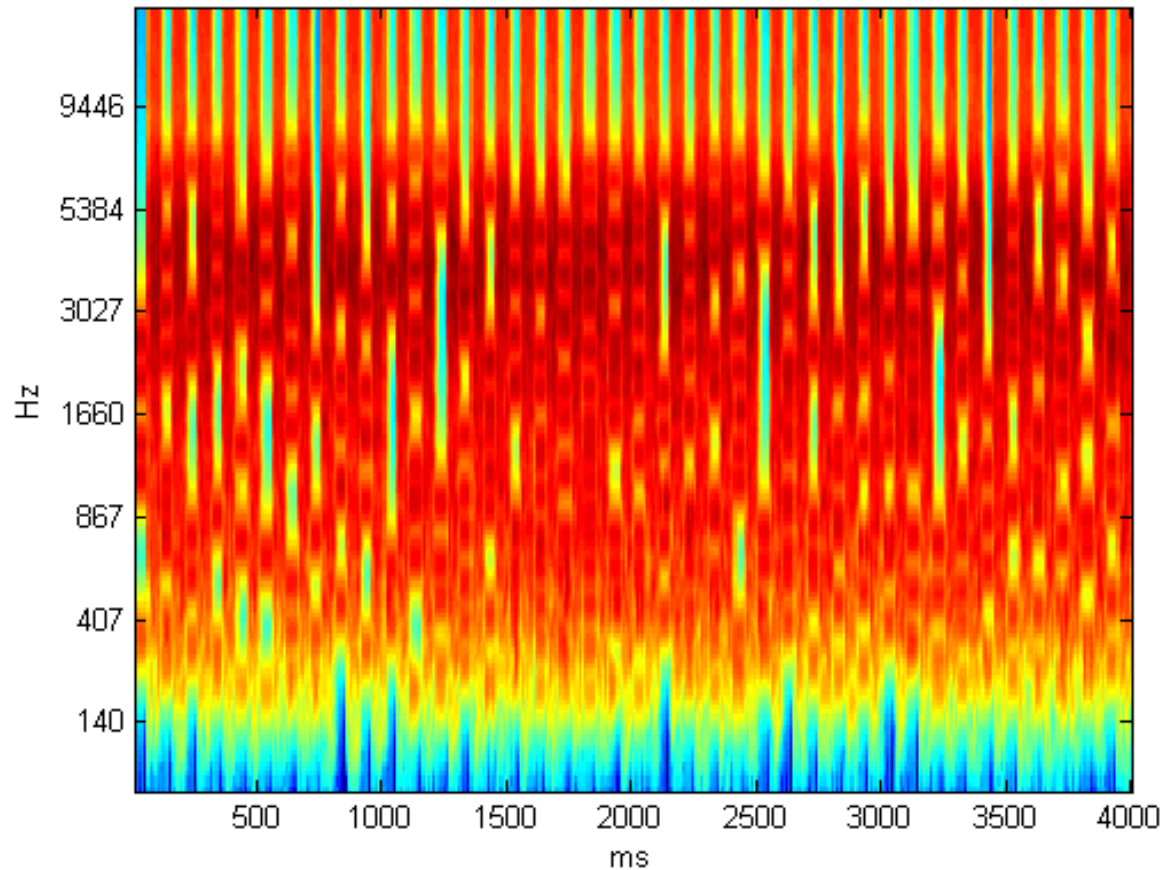


Coherence: [1 2 4 6 8]

Duration: [2:7]

Expt. 3: 'SFG/Noise'

Stimulus: SFG with 50ms chords alternating with 50ms of white noise

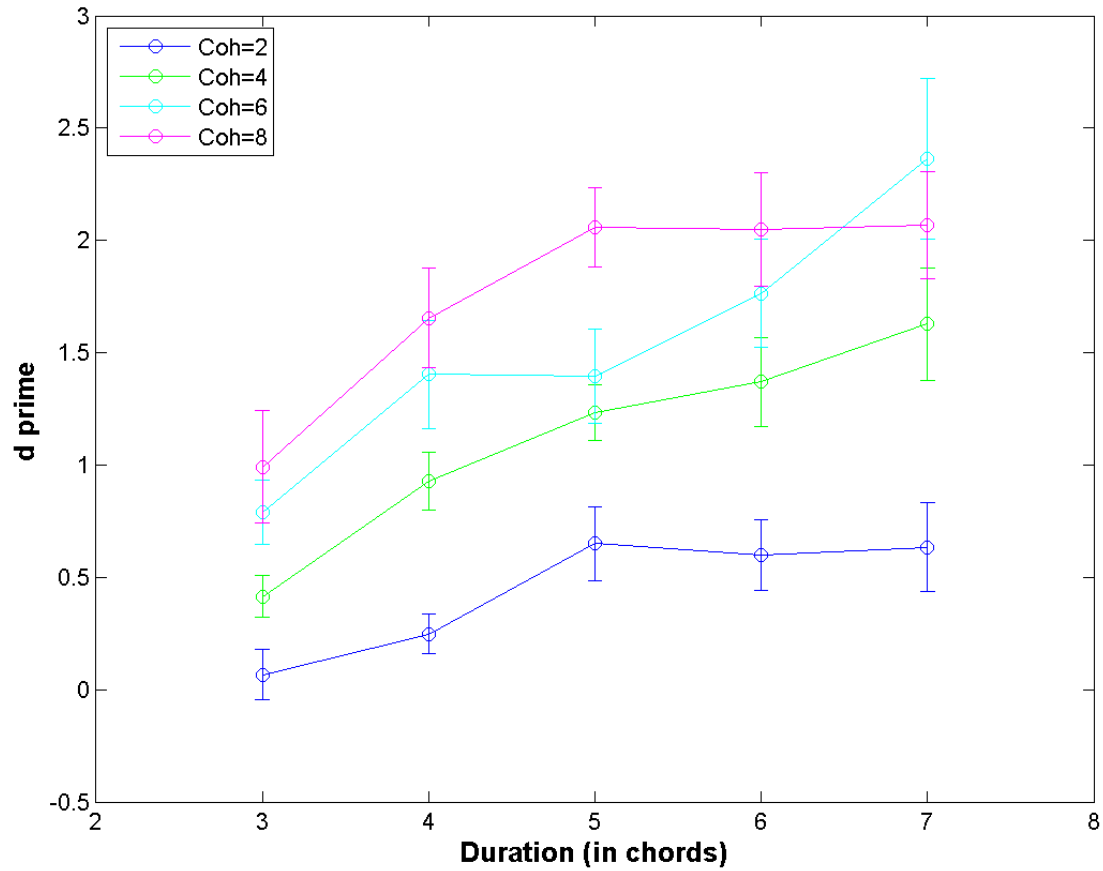


Coherence: [2 4 6 8]

Duration: [3:7]

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

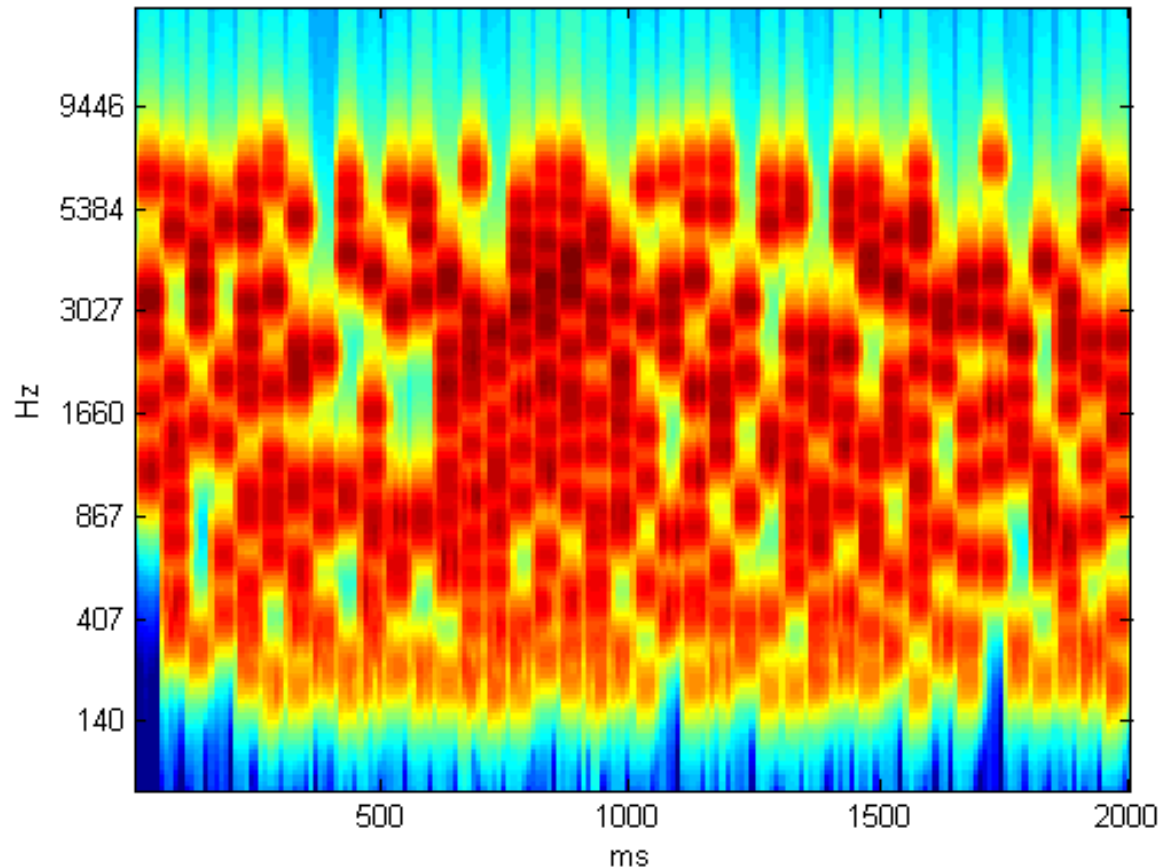
Significant effect of coherence: $F(3, 51) = 23, p < 0.001$

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

No significant effect of condition: $F(1, 17) = 0.004, p = 0.953$

Expt. 4: 'Ramps'

Stimulus: Figures were ramped (successive figure components were not repeating but increasing in frequency in steps of $2l$ or $5l$, where $l = 1/24$ of an octave is the resolution of our frequency pool; ramps within critical band)



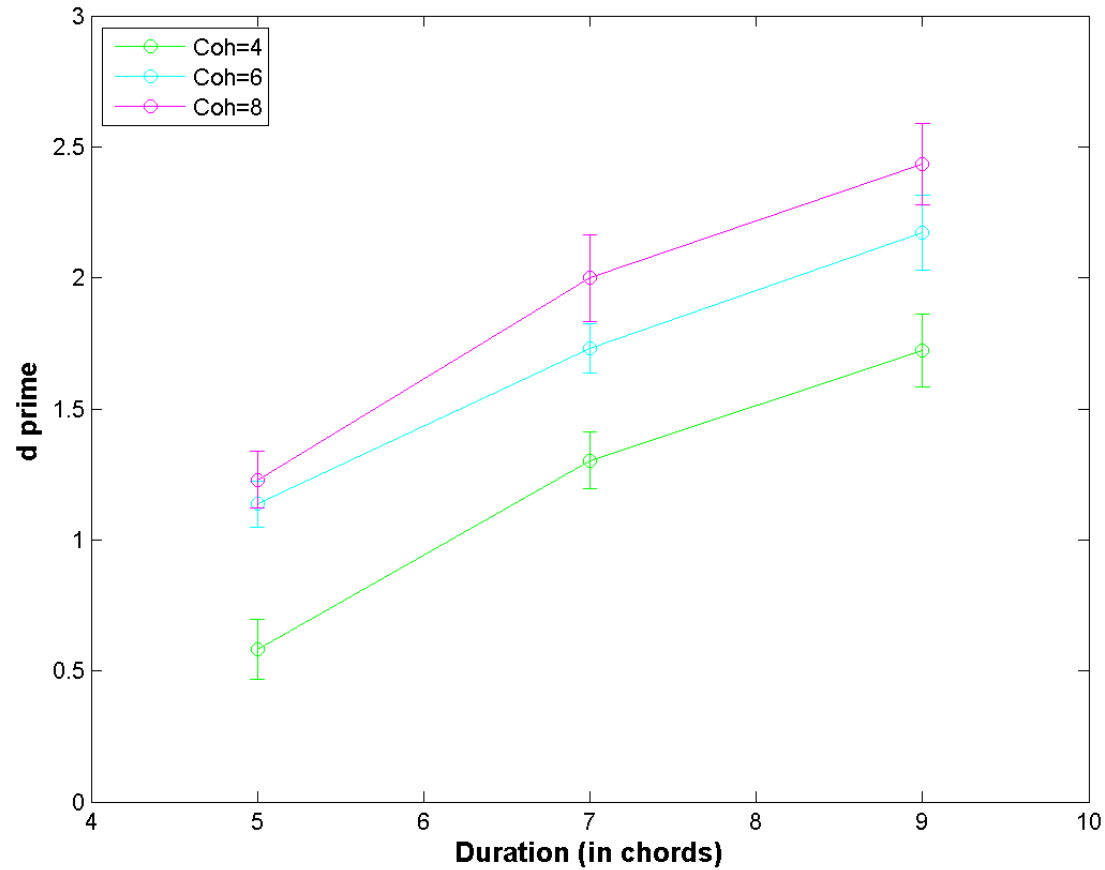
Coherence: [4 6 8]

Duration: [5 7 9]

Ramp step: [2/5]

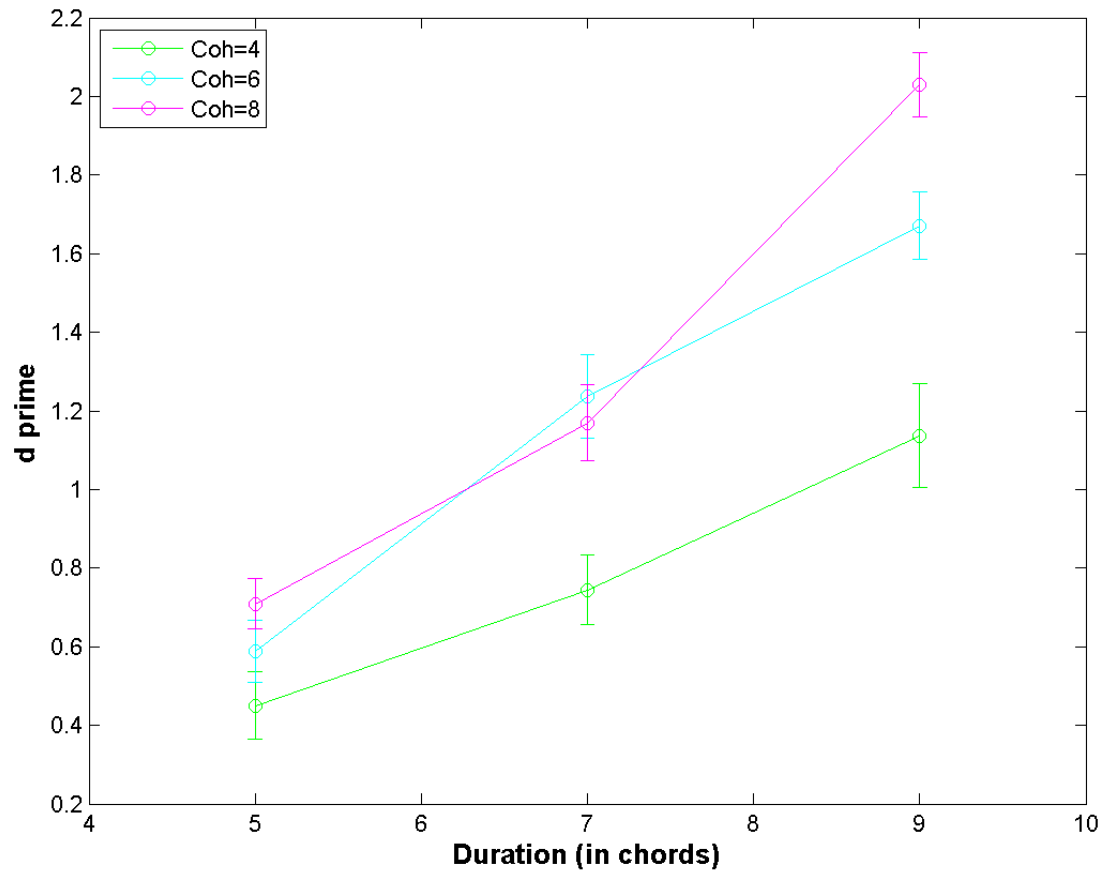
Results: Ramps 2

(n=10)



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

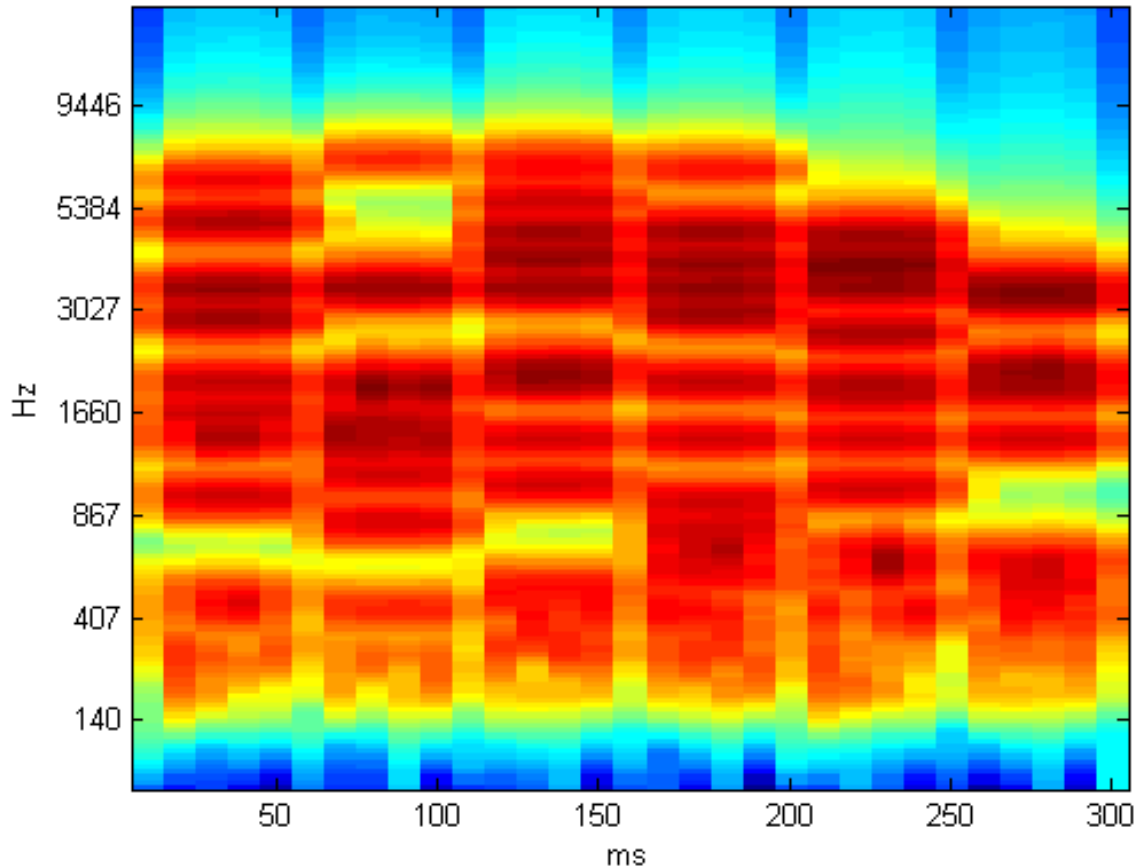


Figure:



Ground:

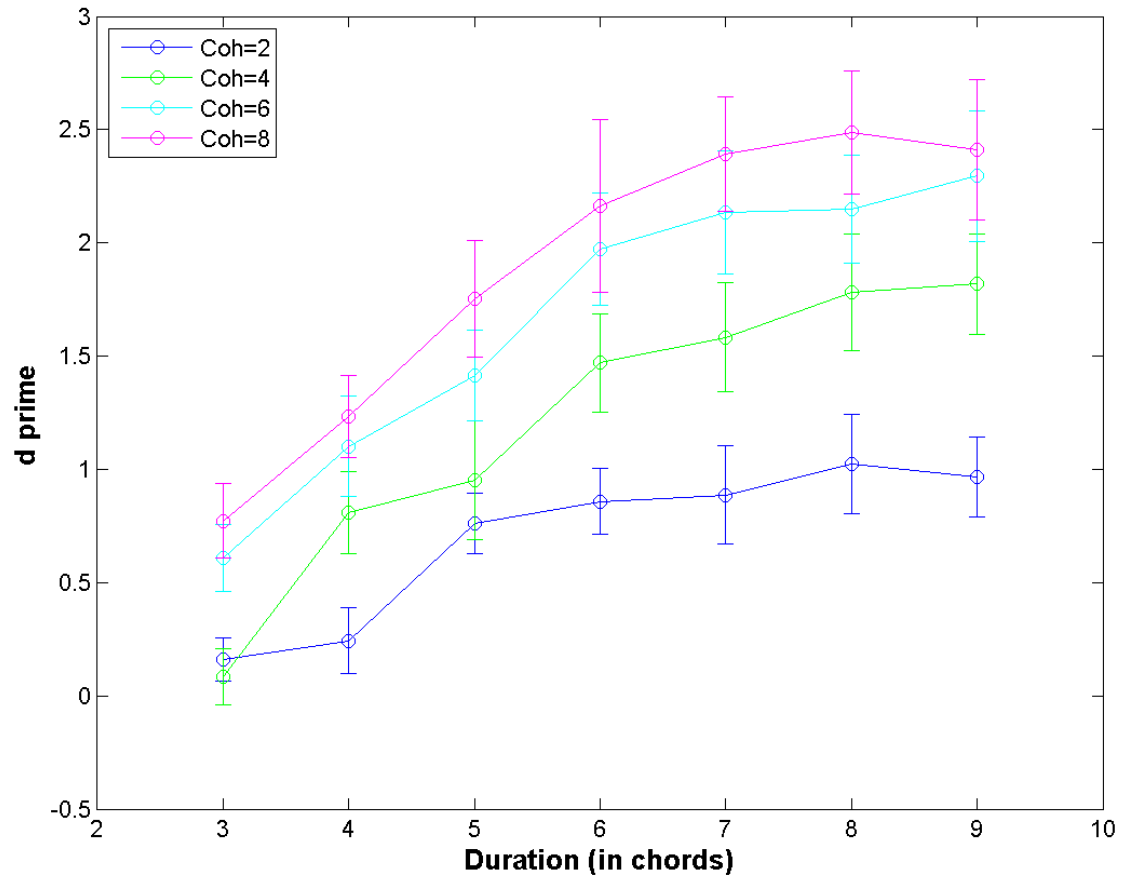


Coherence: [2 4 6 8]

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:

- Invariant to disruption of SFG chords by noise (Expt. 1 & 2)
- Sensitive to shape of figure (continuous vs. ramped) (Expt. 1 & 3)
- Sensitive to size of ramps (2 vs. 5) (Expt. 3a & 3b)
- Invariant to the presence of preceding background (Expt. 1 & 4)

Discussion

- 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 (Shamma et al., 2011).
- ***For modelling results and further analysis, please visit poster #41***

Acknowledgments



Deborah, Sukhbinder and Tim
Newcastle Auditory Group



Maria, Aiysha and Nicolas
UCL Ear Institute

Shihab Shamma
University of Maryland, College Park



Questions ?

Temporal coherence model

The temporal coherence model incorporates three different stages:

- cochlear processing
- processing of the signal at the cortical level
- 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.

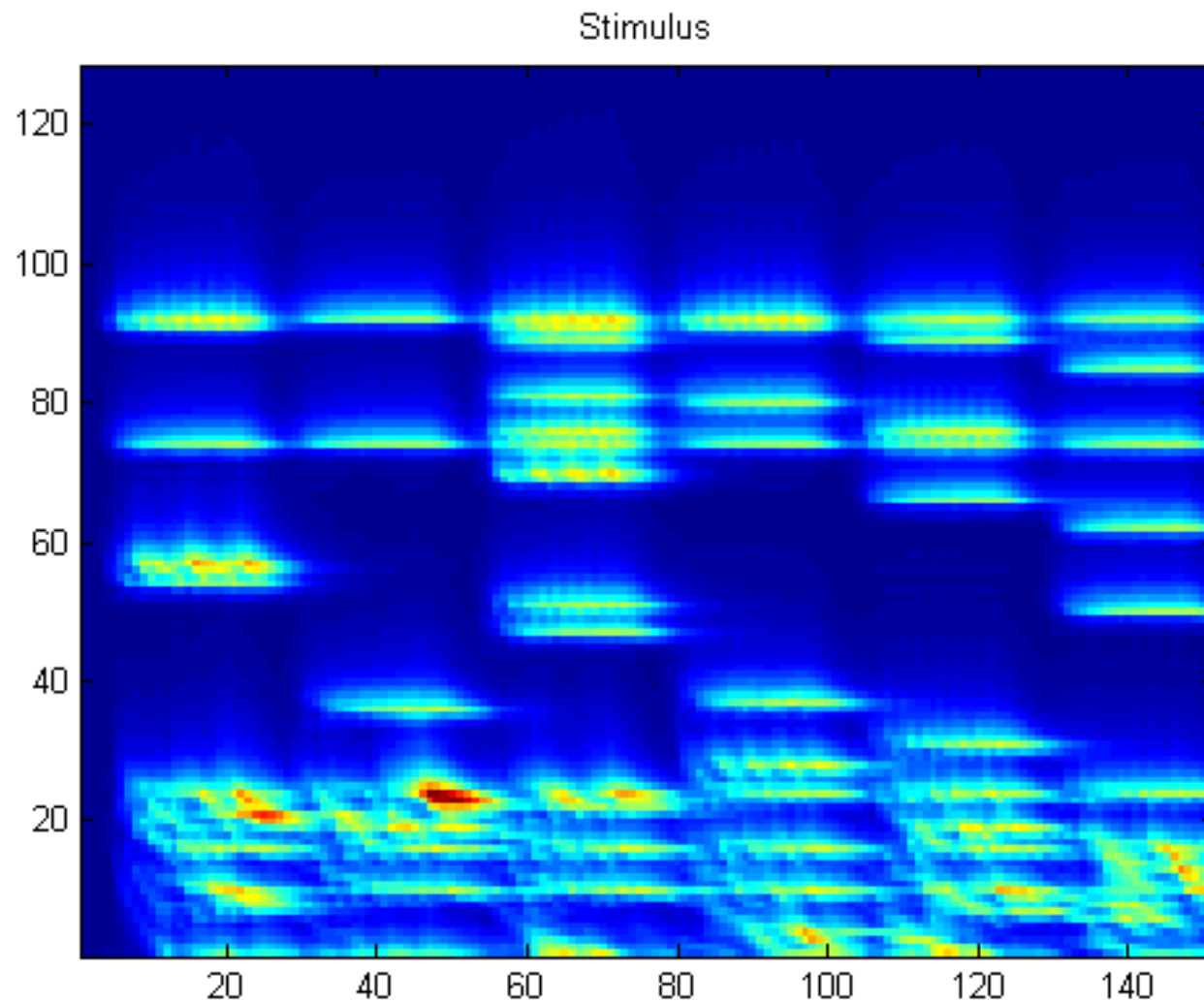
Analysis:

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

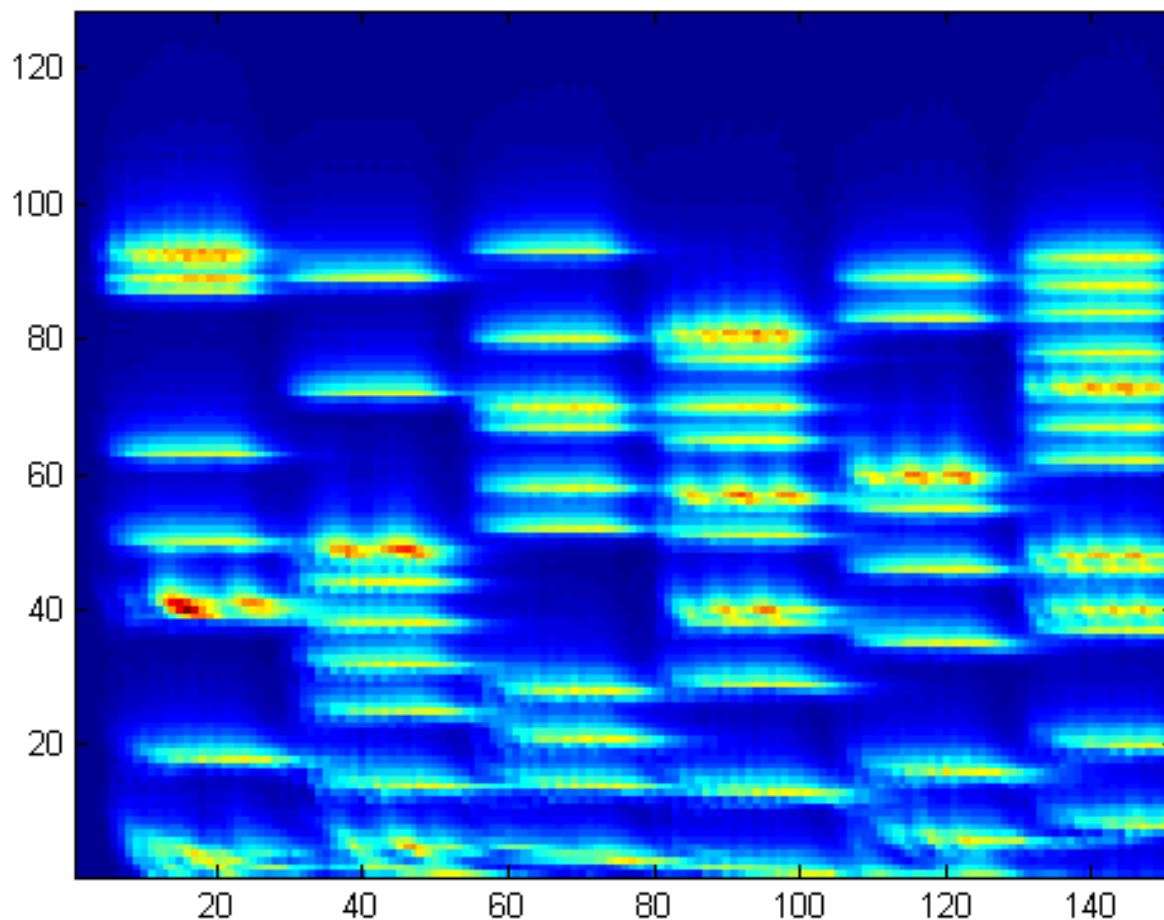
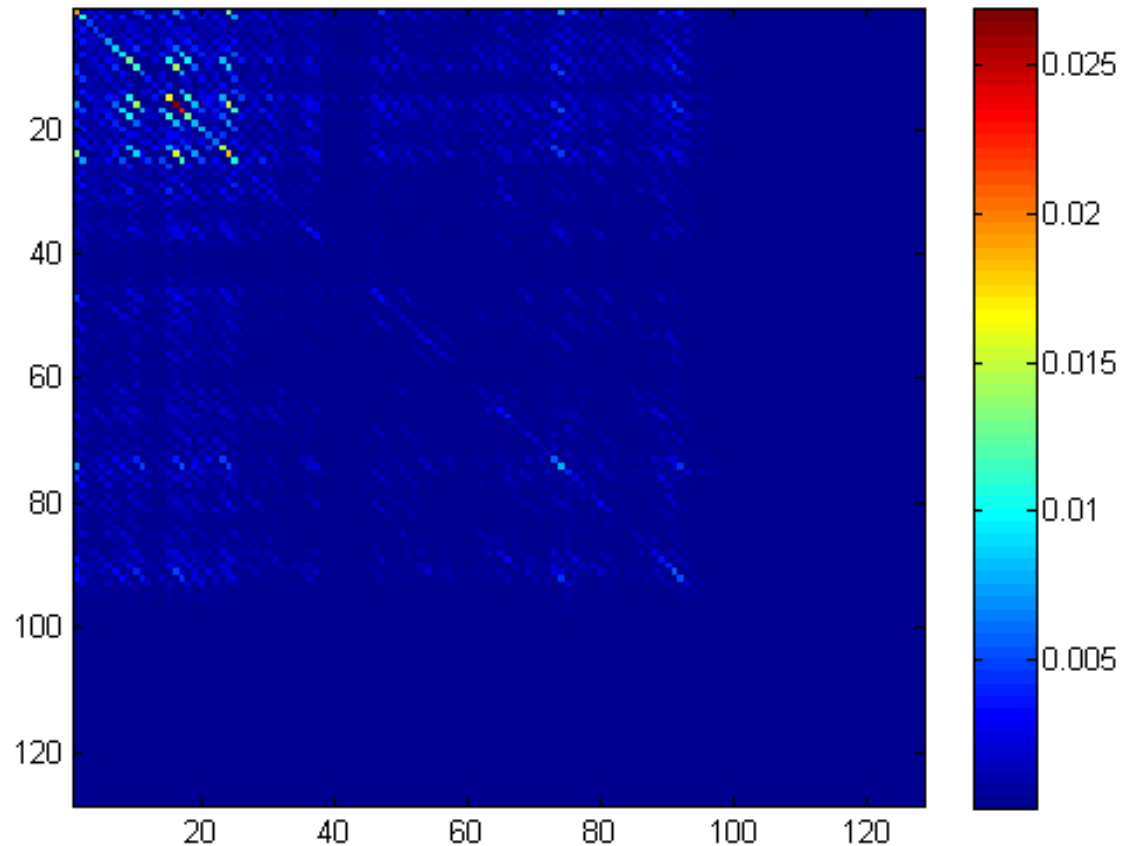
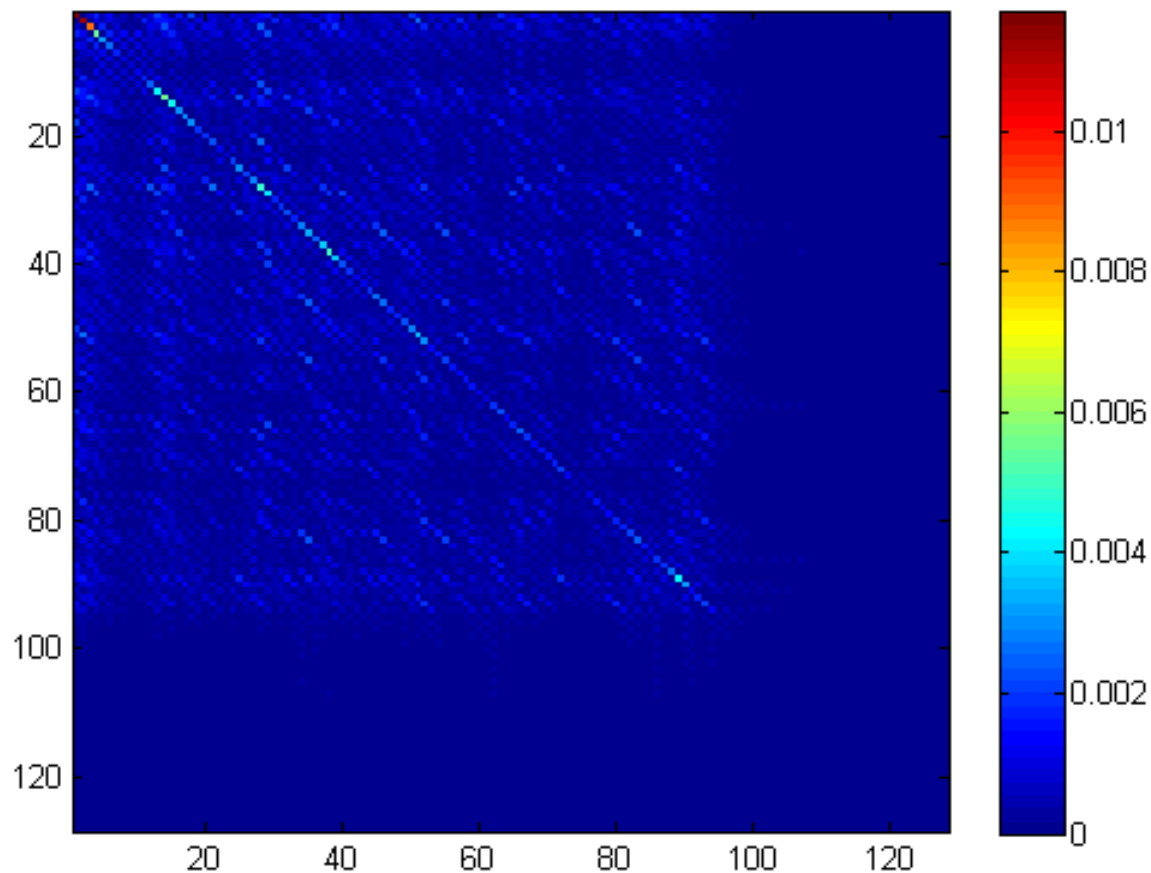


Figure: cross-correlation matrix

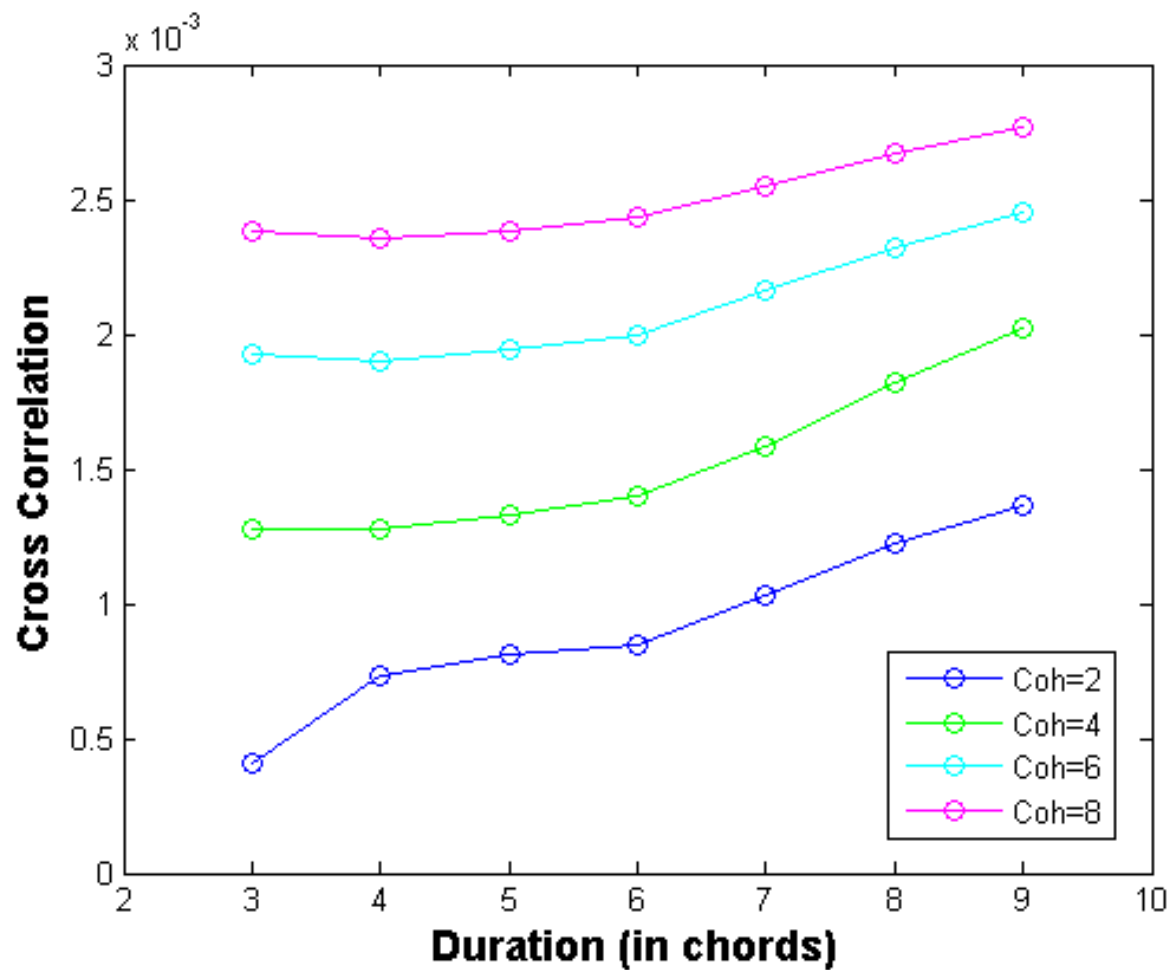


Output: cross-correlation matrix

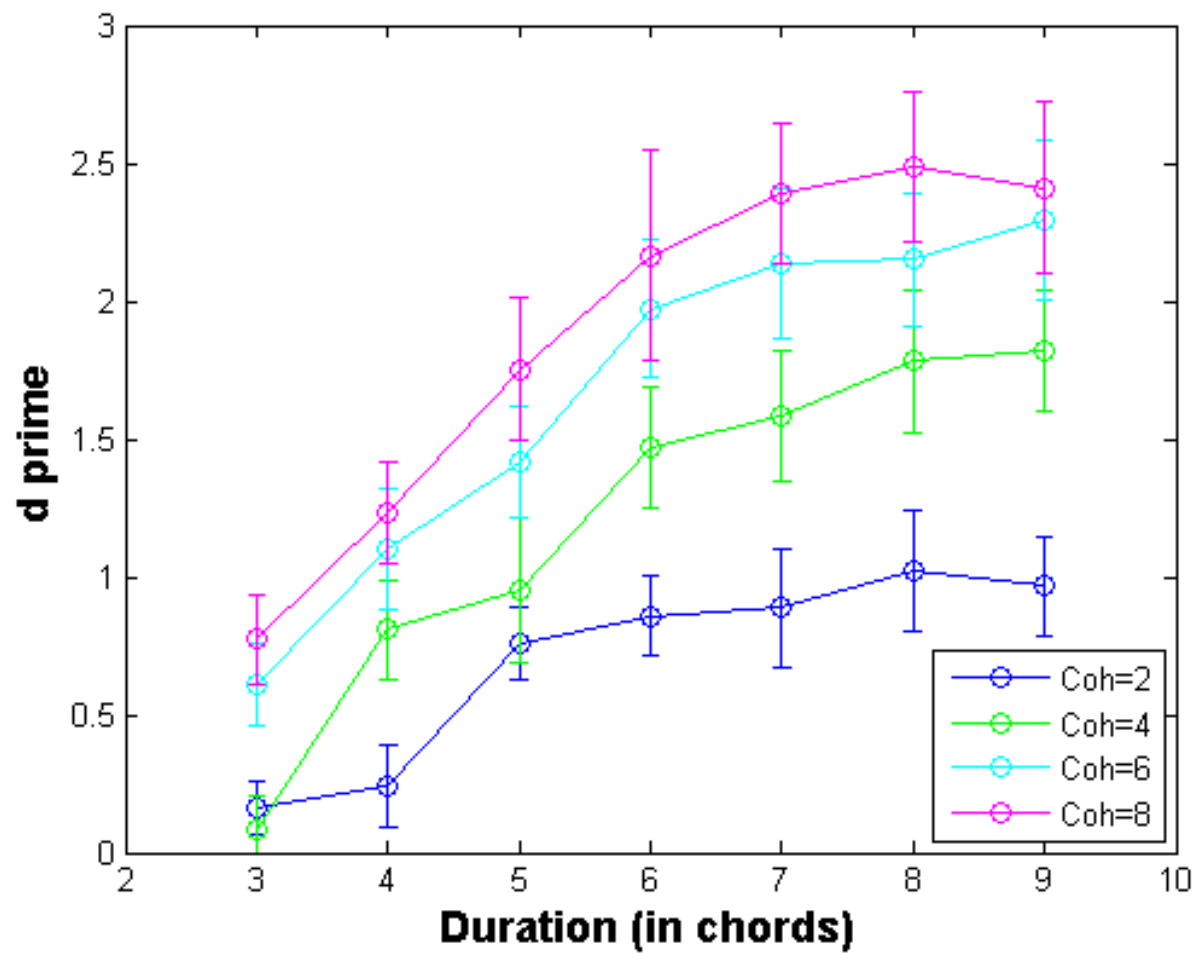
Ground: cross-correlation matrix



Model response



Behavioural response



Acknowledgments



Deborah, Sukhbinder and Tim
Newcastle Auditory Group



Maria, Aiysha and Nicolas
UCL Ear Institute

Shihab Shamma
University of Maryland, College Park

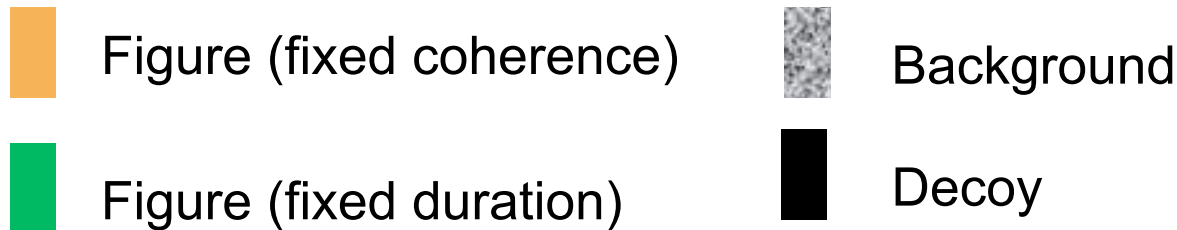
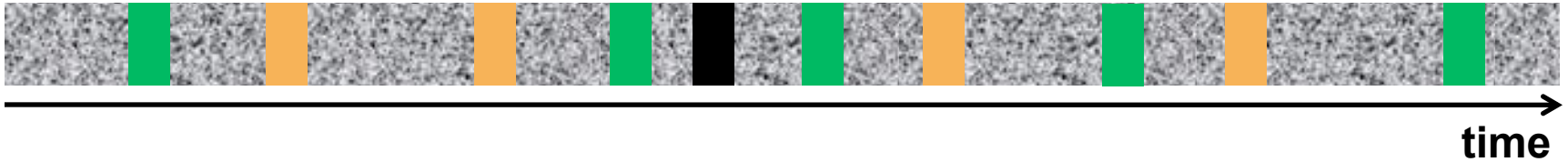


fMRI Experiment

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

Stimulus:

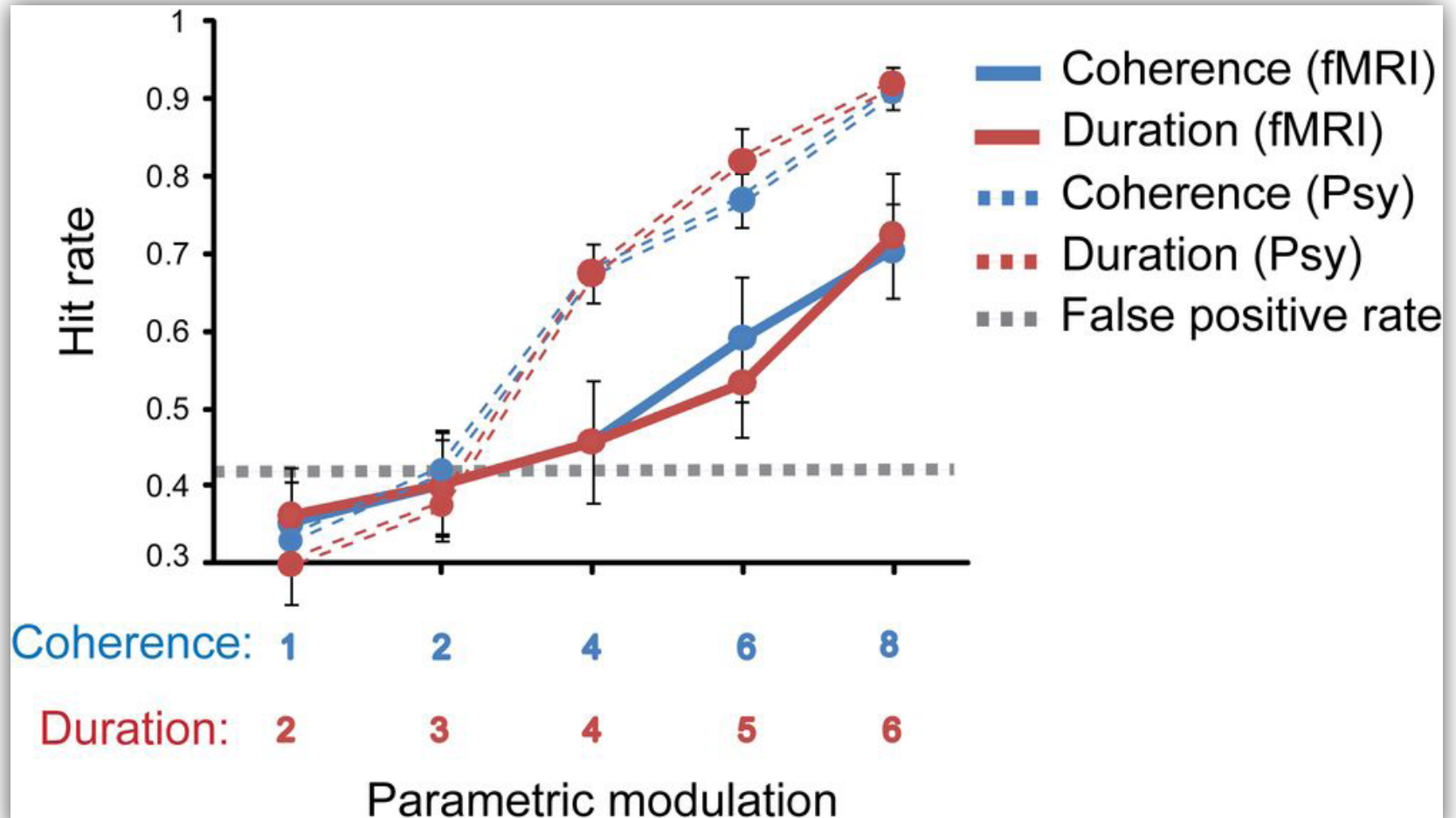
- i. Fixed coherence: 4, varying duration: 2-7 chords
- ii. Fixed duration: 4, varying coherence: 1,2,4,6,8



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

➤ **Subjects were not actively detecting figures**

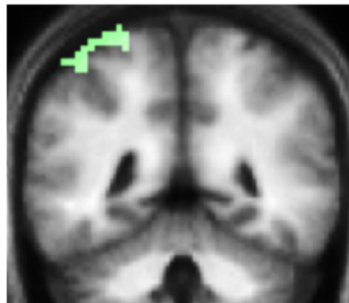
Behaviour in scanner



Effects of Duration

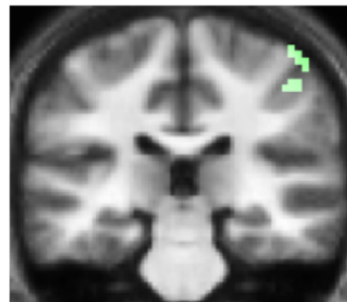
A

Left IPS

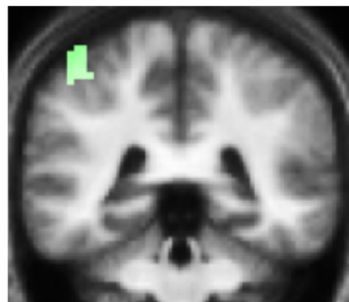


y = -46

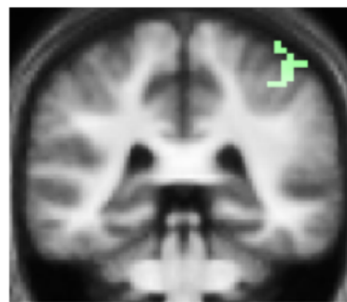
Right IPS



y = -28



y = -40



y = -37

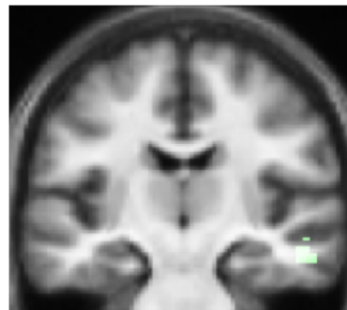
B

Left STS



y = -34

Right STS

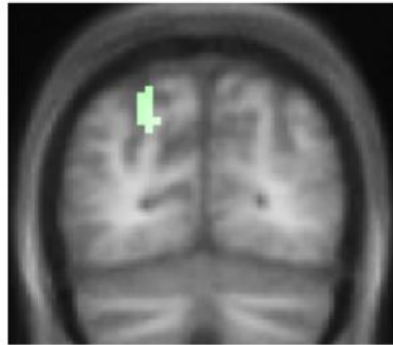


y = -13

Effects of Coherence

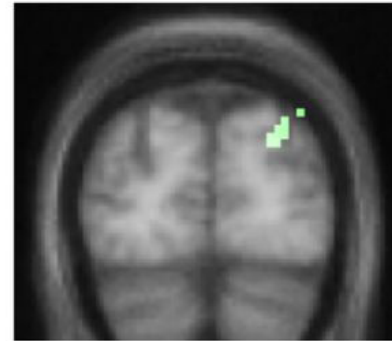
A

Left IPS



$y = -73$

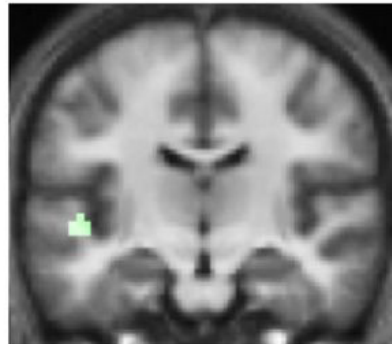
Right IPS



$y = -82$

B

Left STS



$y = -16$

Right STS



$y = -4$