Auditory figure-ground segregation in complex acoustic scenes

Sundeep Teki [1]

Maria Chait [2]

[1] Wellcome Trust Centre for Neuroimaging, University College London, UK
[2] UCL Ear Institute, University College London, UK

Outline

- Introduction
- Figure-ground stimulus
- fMRI study
- Psychophysics
- Temporal coherence model
- Summary

The problem

 # speakers

1 

2 

4 

8 

Auditory figure-ground segregation

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

Processes:

- grouping of simultaneous frequency components
- grouping of frequency components over time,
- separation of grouped components from the rest of the acoustic scene.

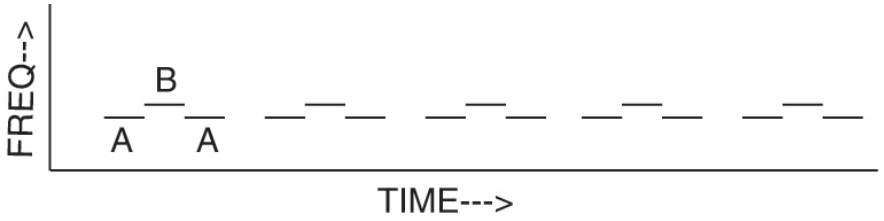
Stimuli:

- Studied using relatively simple signals, e.g. streaming signals

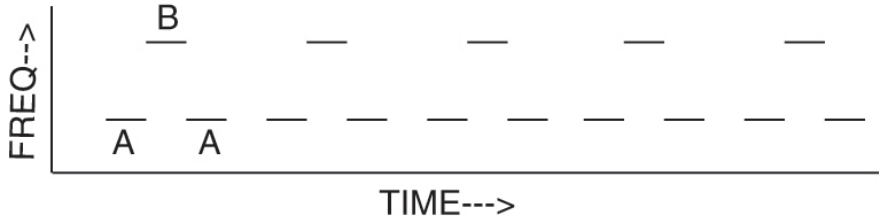
(van Noorden, 1975; Bregman, 1990)

Streaming

ABA_…ABA_ or AB…AB

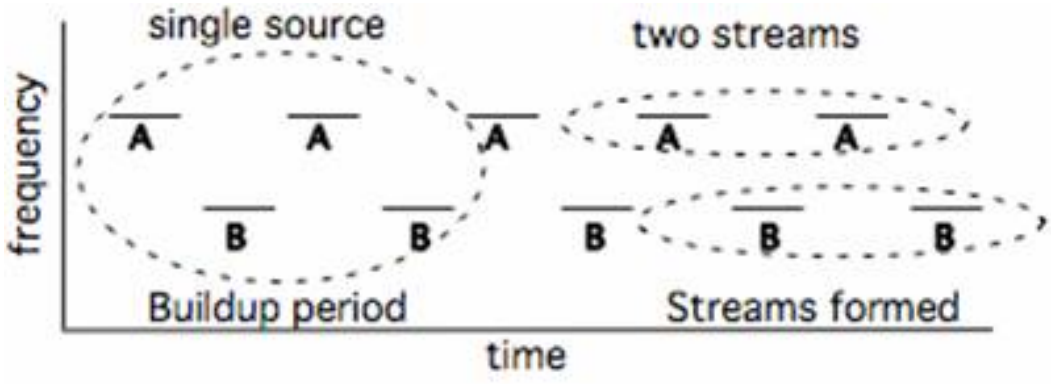


Δf: 1 st **Integrated percept**



Δf: 6 st **Ambiguous percept**

Δf: 9 st **Segregated percept**



Auditory figure-ground segregation

Mechanisms:

- frequency selectivity
- forward suppression
- neural adaptation

cf. Fishman & Steinchneider; Bee & Klump; Micheyl; Carlyon

Auditory figure-ground segregation

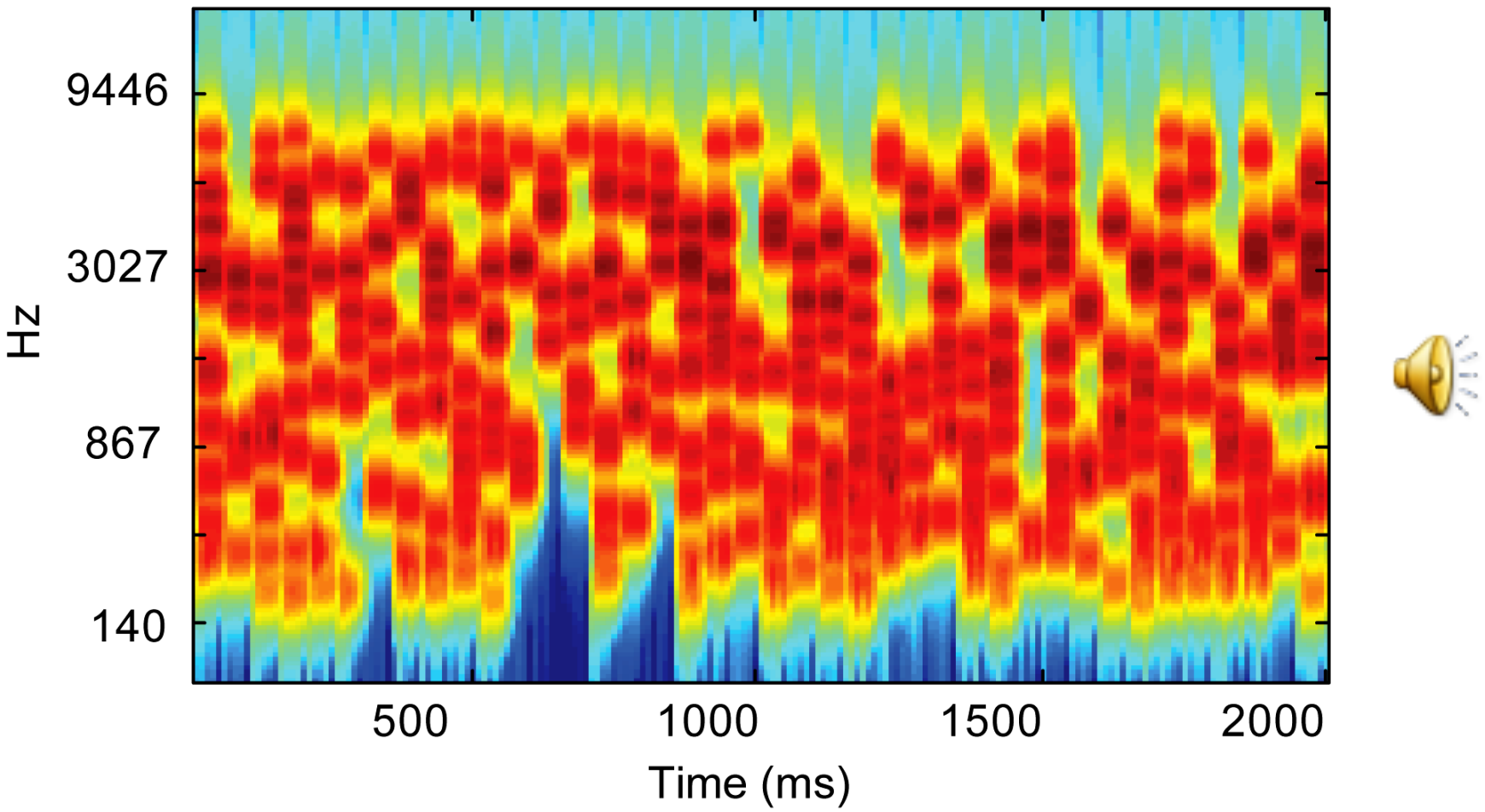
Drawbacks of streaming signals:

- lack the rich spectrotemporal complexity of natural signals
- predictable temporal structure
- spectral components do not change with time

Outline

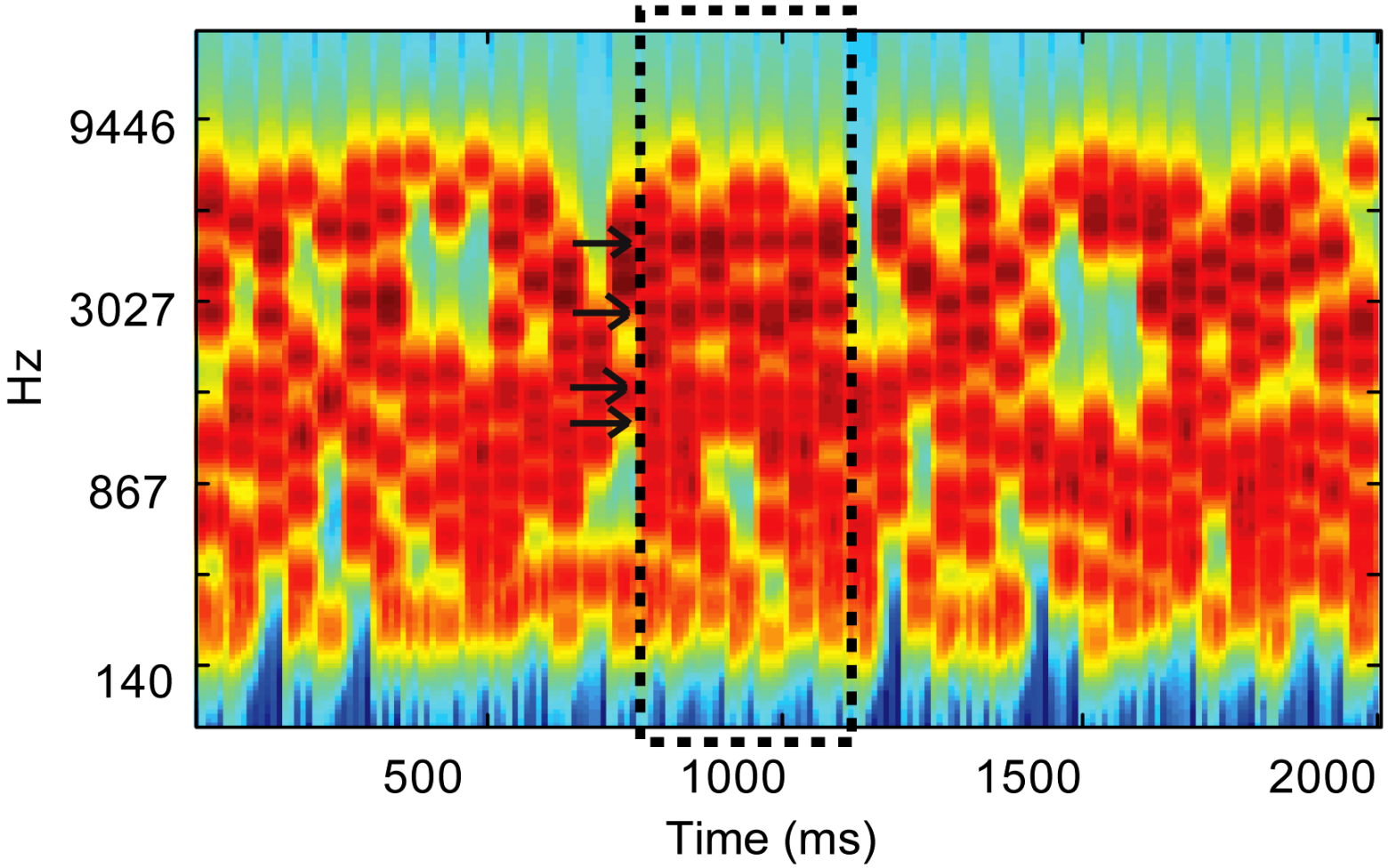
- Introduction
- **Figure-ground stimulus**
- **fMRI study**
- **Psychophysics**
- **Temporal coherence model**
- **Summary**

Stochastic Figure-Ground (SFG)



SFG: Figure present

Figure with ‘coherence’ = 4 and ‘duration’ = 7



SFG: Stimulus design

Stimulus:

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
Resolution of frequency pool:	1/24th of an octave
Cosine ramp:	10 ms for onset and offset

Coherence:

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

Duration:

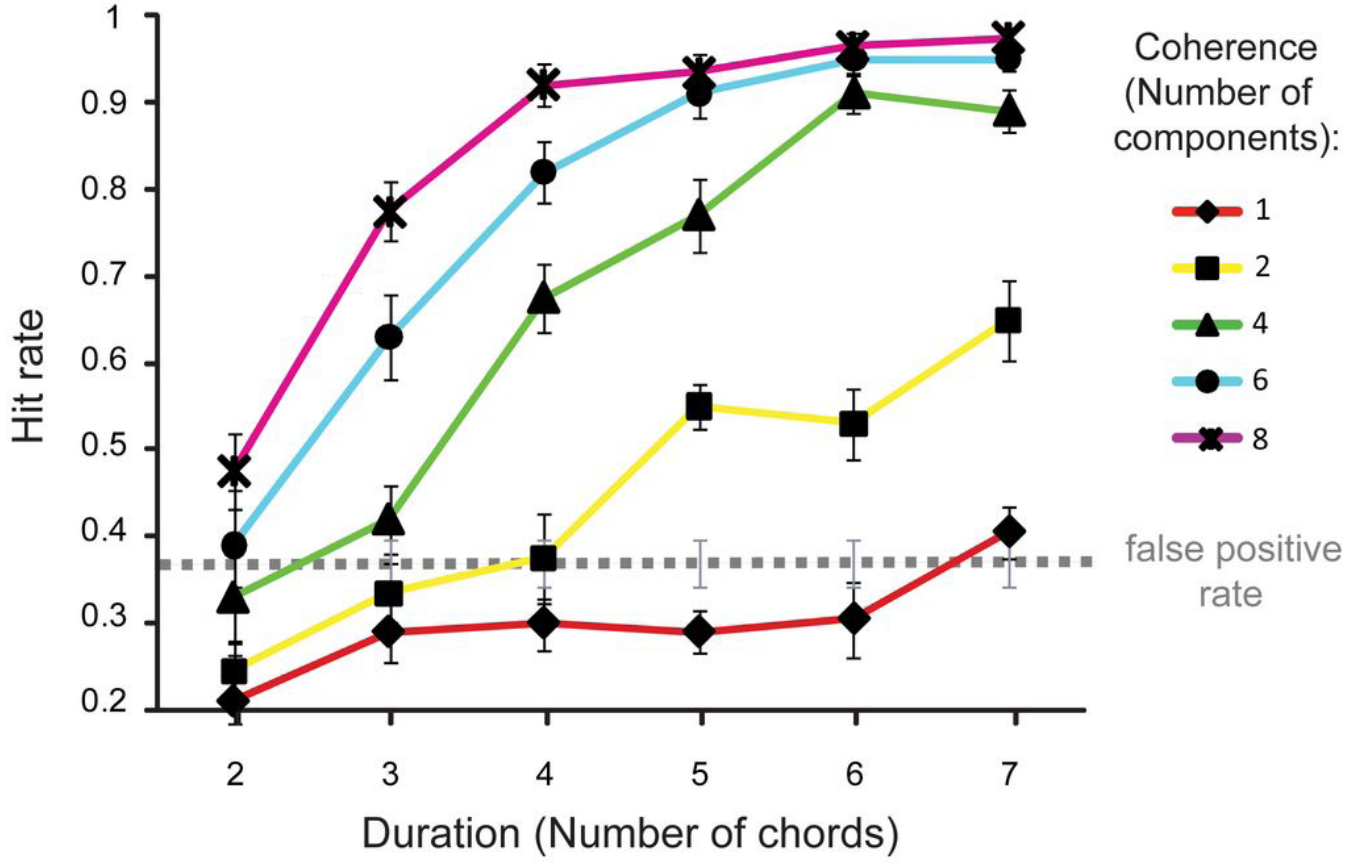
Number of chords over which frequencies repeat : **2-7**

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

Psychophysics

n=10



- Listeners are remarkably sensitive to the appearance of figures
- Sensitive to parametric variations of coherence and duration

Outline

- Introduction
- Figure-ground stimulus
- **fMRI study**
- **Psychophysics**
- **Temporal coherence model**
- **Summary**

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

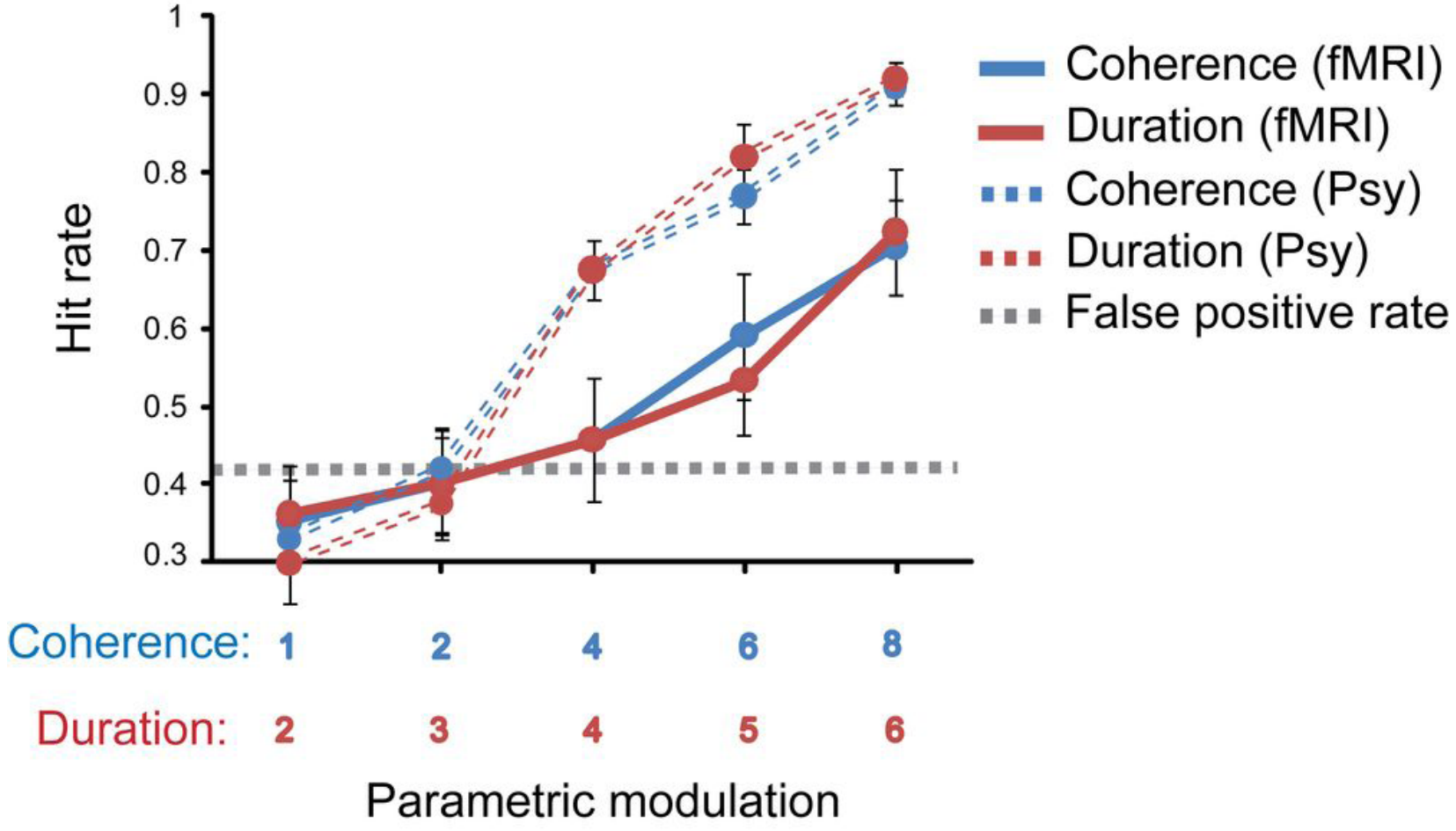
Paradigm:

i. Passive listening

ii. Active detection

- 3 Tesla Siemens Allegra MRI Scanner
- Continuous scanning
- 42 contiguous slices per volume
- TR: 2.52 s; TA: 2.88 s; TE: 30 ms
- Slice thickness: 2 mm with 1mm gap between slices
- In-plane resolution: 3.0 x 3.0 mm^2
- 3 scanning sessions: 510 volumes per subject

Behaviour in scanner



fMRI design

Figure (fixed coherence)

Figure (fixed duration)

Background

Decoy

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

- **Subjects were not actively detecting figures**

fMRI analysis

- 14 subjects (normal hearing, no audiological disorders)
- Standard pre-processing with SPM8
- Whole brain analysis
- Statistical model based on General Linear Model
- Random effects design

Parametric Modulation:

I. Effect of Duration: Fixed coherence (4); varying duration (2-7)

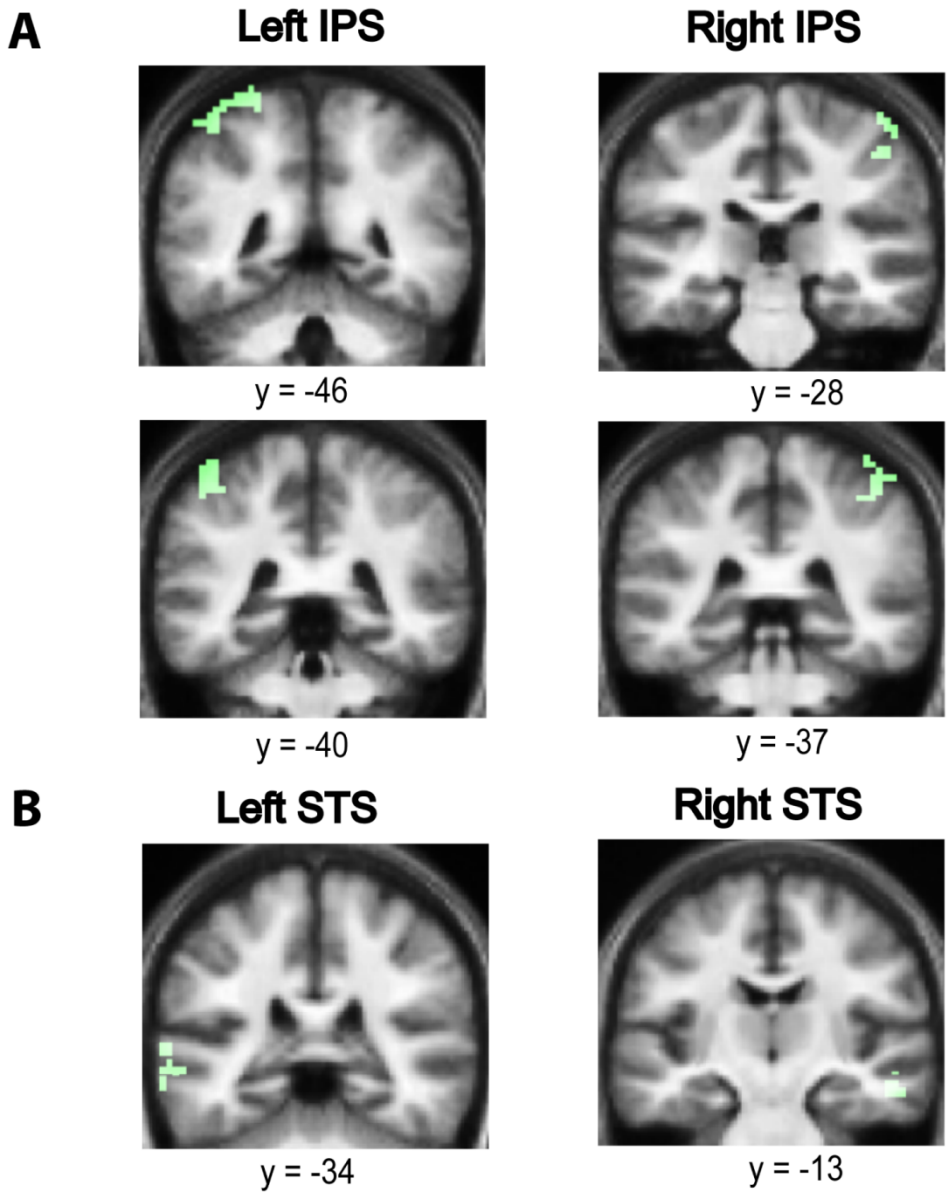
II. Effect of Coherence: Fixed duration (4); varying coherence (1,2,4,6,8)

fMRI Results

I. Effects of Duration:

Intraparietal Sulcus (IPS)	(bilateral; anterior)
Superior Temporal Sulcus (STS)	(bilateral)
Planum Temporale	(R)
Medial Geniculate Body (MGB)	(bilateral)

Effects of Duration



fMRI Results

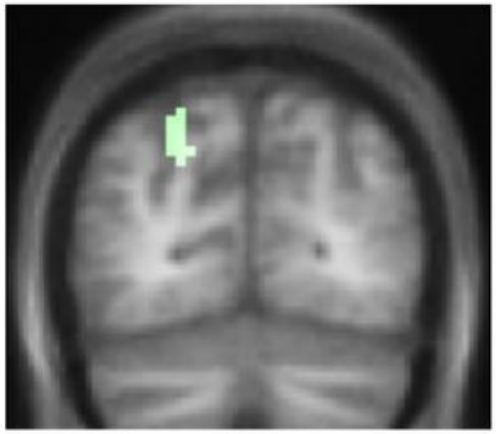
II. Effects of Coherence:

Intraparietal Sulcus	(bilateral; posterior)
Superior Temporal Sulcus	(bilateral)

Effects of Coherence

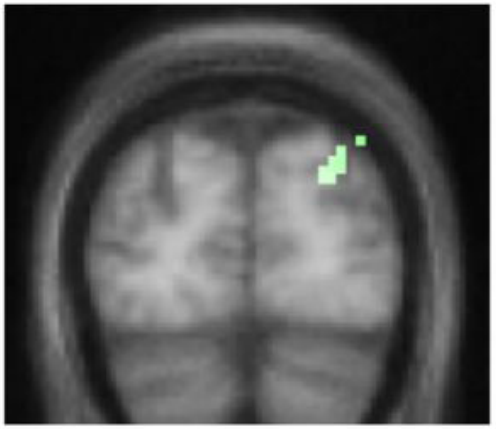
A

Left IPS



y = -73

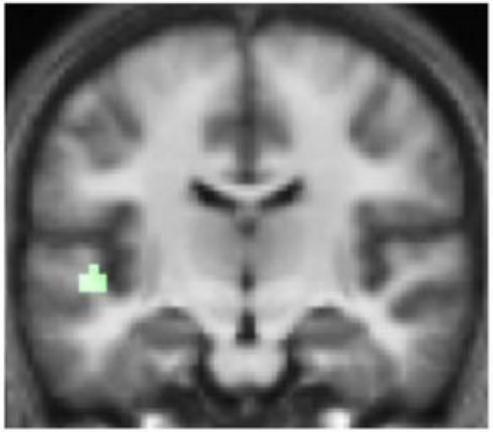
Right IPS



y = -82

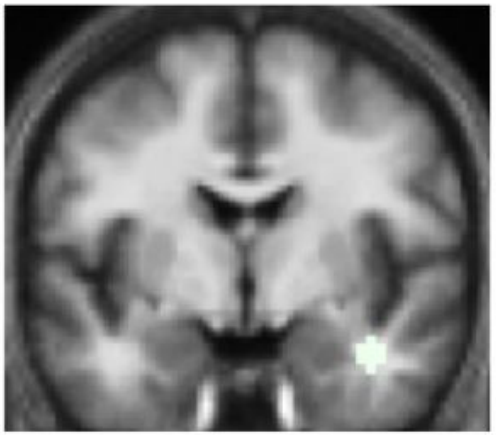
B

Left STS



y = -16

Right STS



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 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***

fMRI 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?
- Is IPS causally responsible for segregation?

Teki, Chait et al., J Neurosci (2011)

Outline

- Introduction
- Figure-ground stimulus
- fMRI study
- **Psychophysics**
- **Temporal coherence model**
- **Summary**

Psychophysics

Aims:

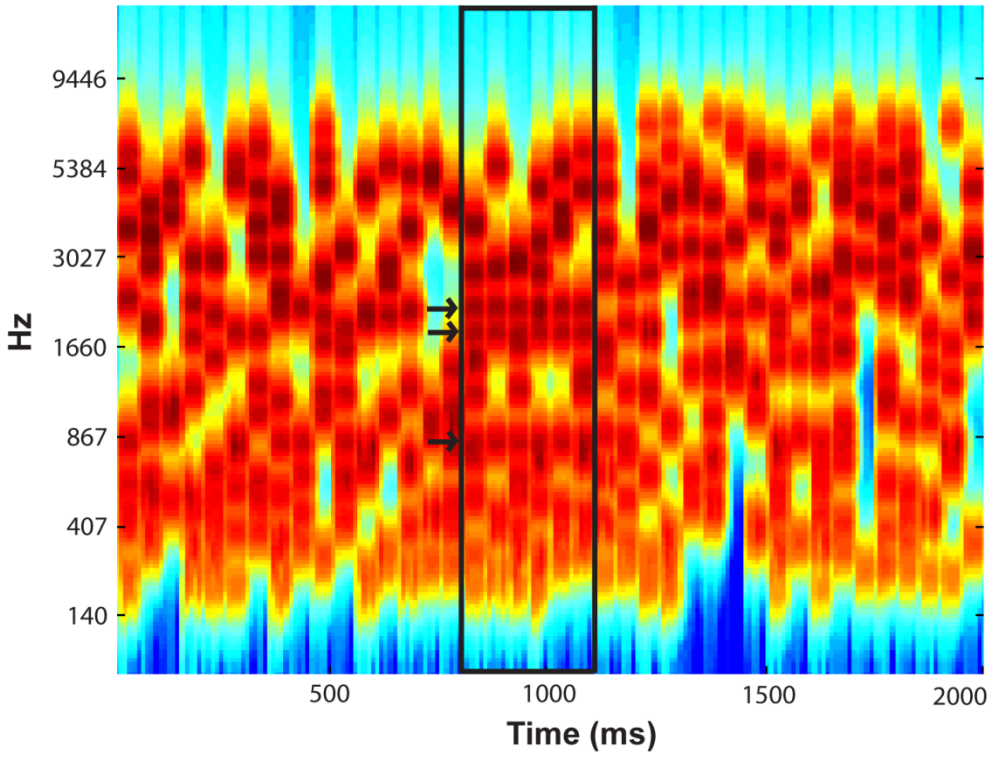
- To characterize the brain mechanisms that underlie complex figure-ground segregation through systematic manipulations of the SFG stimulus

- To examine sensitivity to figures by introducing systematic perturbations

- Test role of adaptation in mediating segregation in our complex stimulus

Expt. 1: ‘Baseline’ (50 ms)

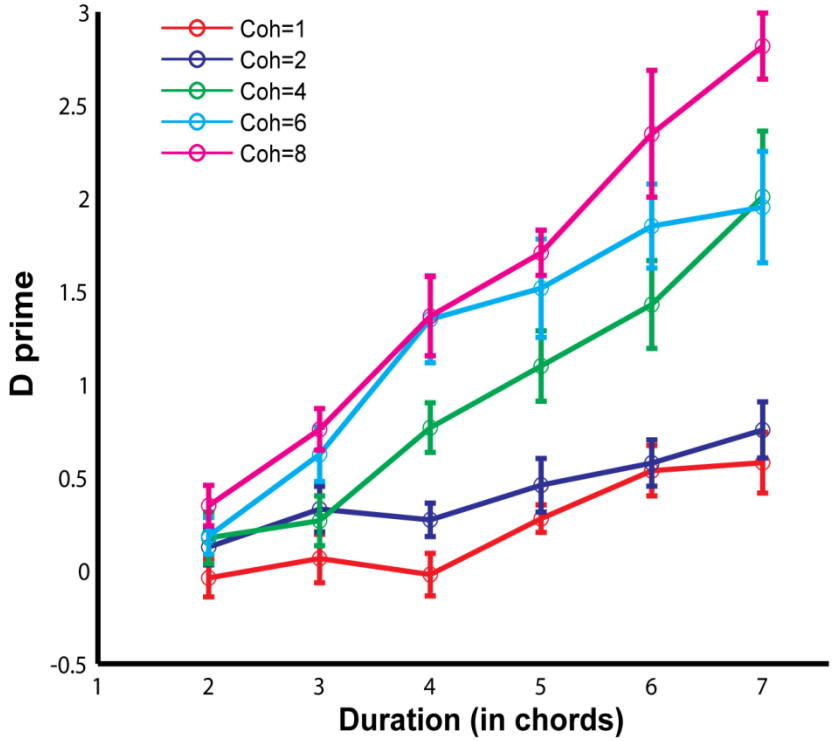
Stimulus consisted of a sequence of 40 x 50ms chords (2 s long)



Coherence: [1 2 4 6 8] Duration: [2:7]

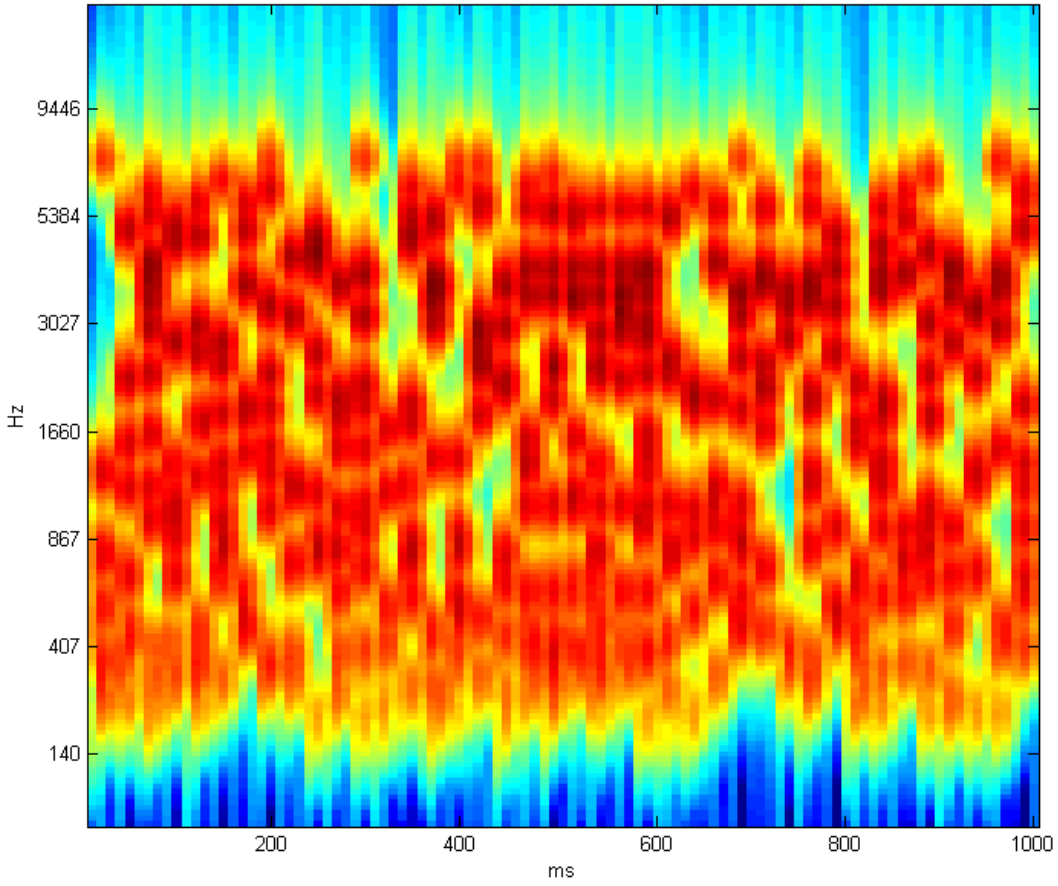
Expt. 1: Results

(n=9)



Expt. 2: ‘Baseline’ (25 ms)

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

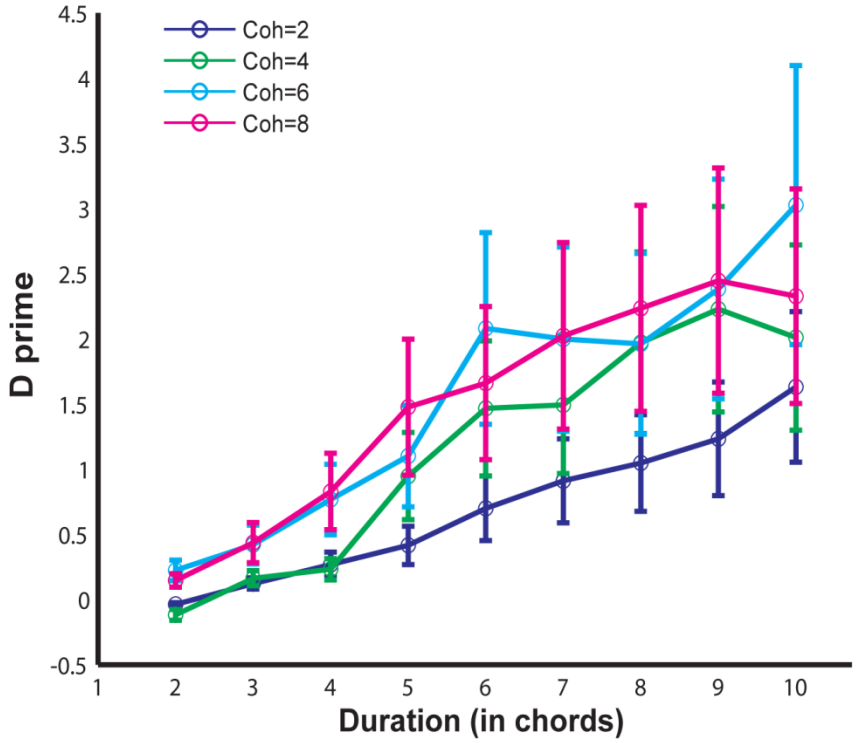


Coherence: [2 4 6 8]

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

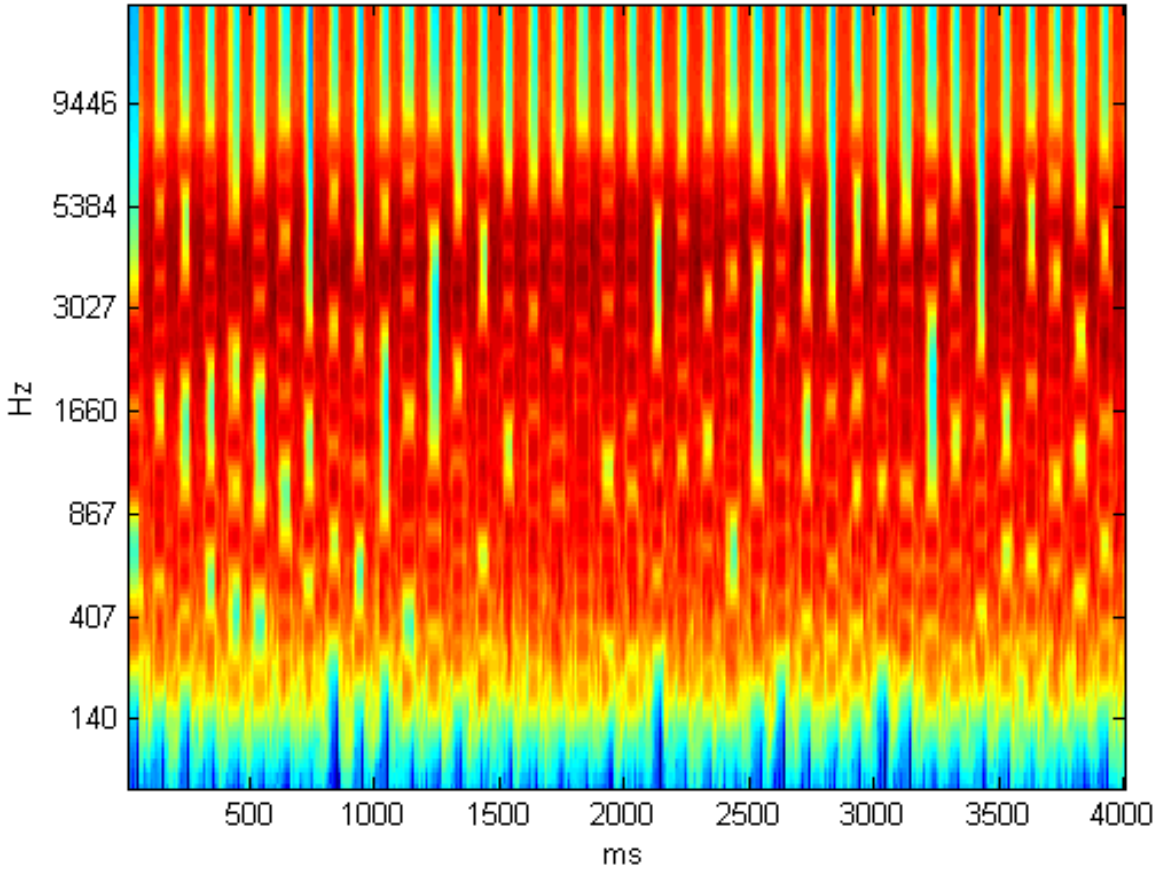
Significant effect of coherence: $F(3, 45) = 77, p < 0.001$

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)

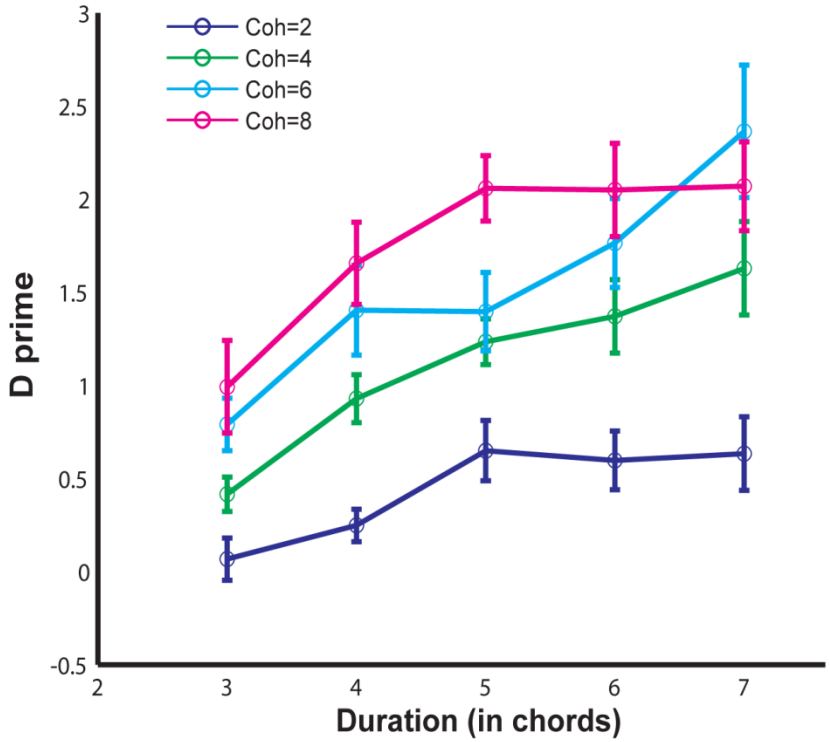


Coherence: [2 4 6 8]

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

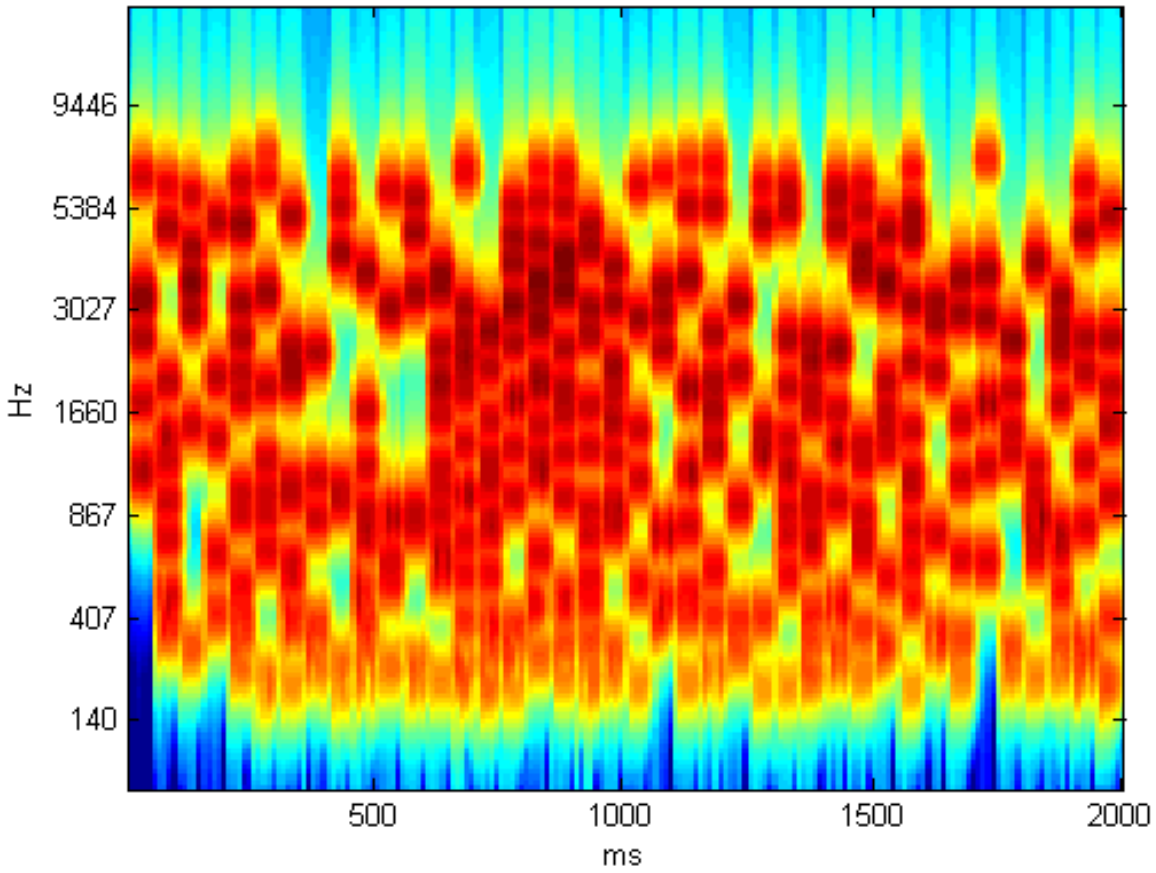
Significant effect of coherence: $F(3, 51) = 23, p < 0.001$

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’

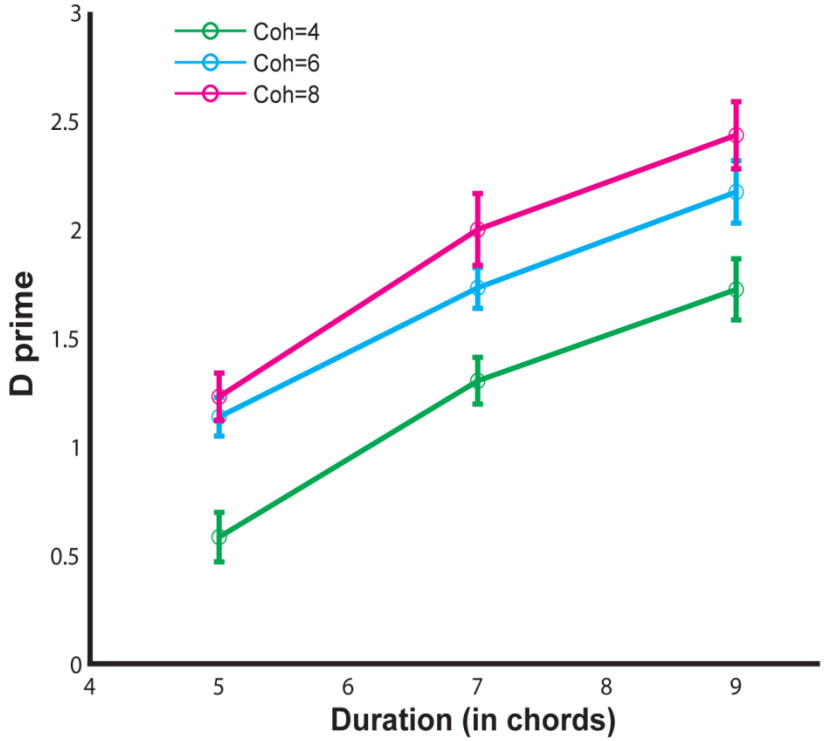
Stimulus: Figures were ramped (successive figure components were not repeating but increasing in frequency in steps of 2*l* or 5*l*, where *l* = 1/24 of an octave is the resolution of our frequency pool; ramps within critical band)



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

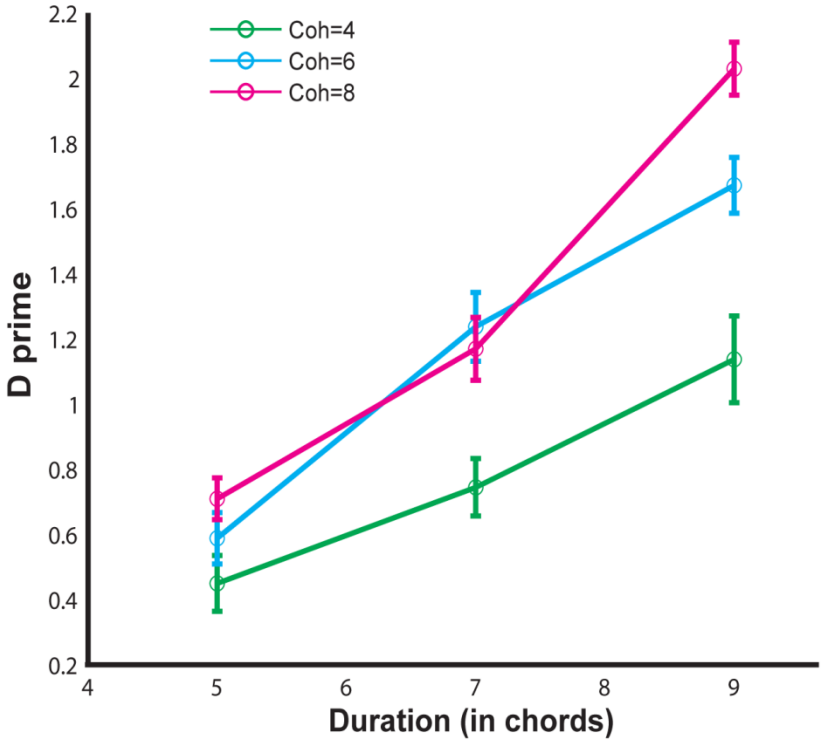
Results: Ramps 2

(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

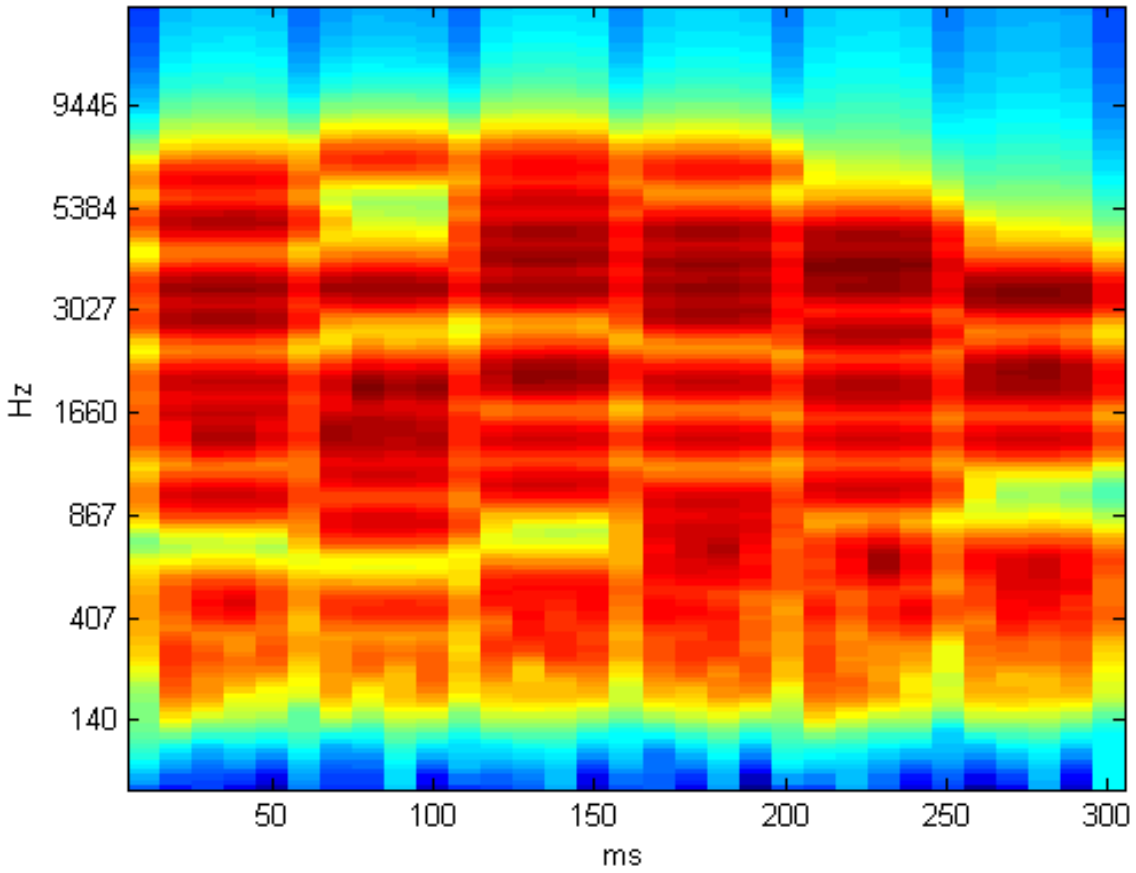


Figure:

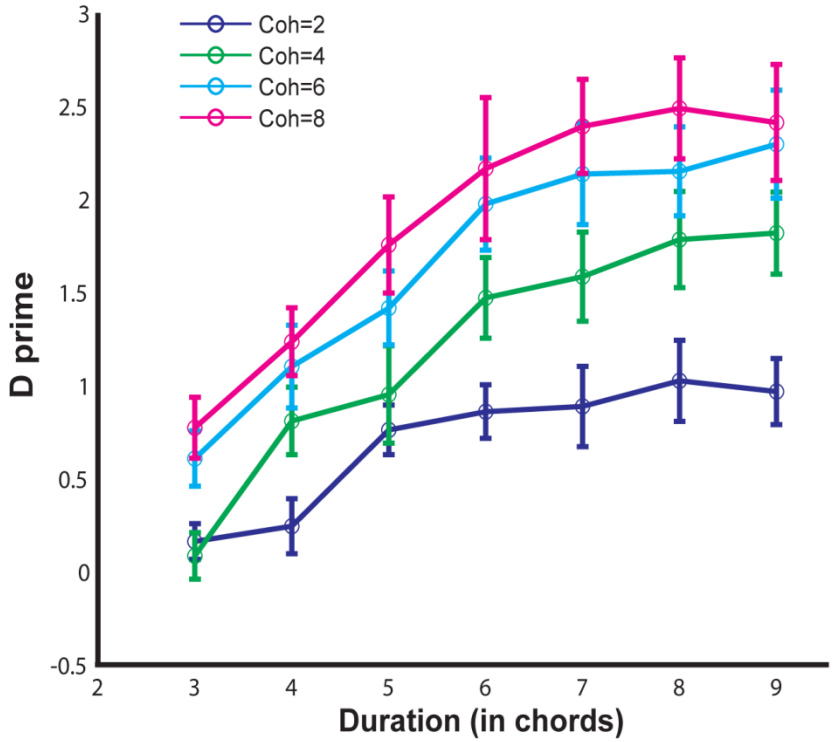
Ground:

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: $F(1,16) = 0.033, p = 0.859$

Psychophysics summary

Figure-detection performance in complex SFG stimulus is:

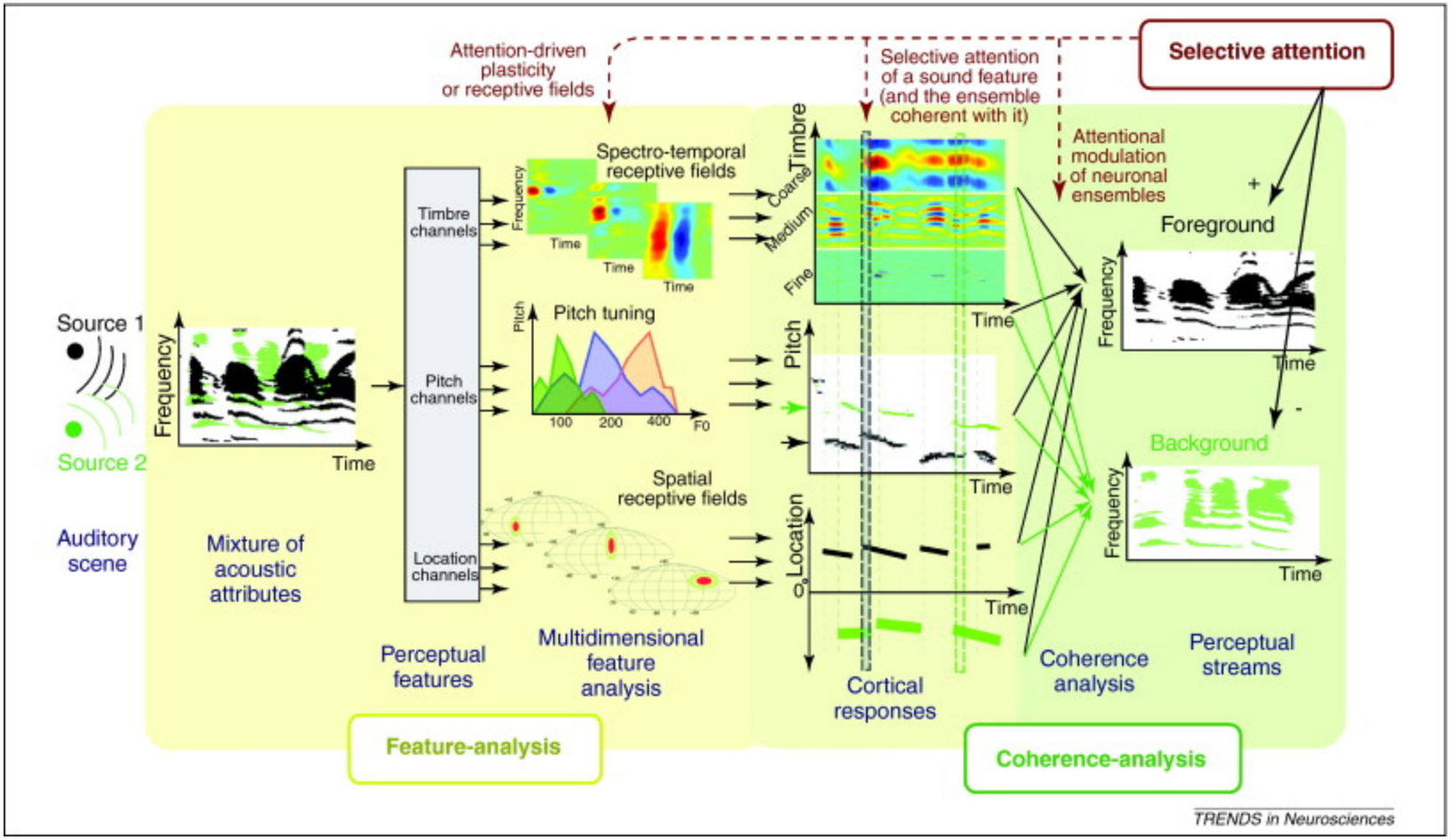
- Depends on no. of repeating chords, not duration of figure (Expt. 1 & 2)
- Invariant to disruption by white noise (Expt. 1 & 3)
- Sensitive to shape of figure (continuous vs. ramped) (Expt. 1 & 4)
- Sensitive to size of ramps (2 vs. 5) (Expt. 4a & 4b)
- Invariant to the presence of preceding background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

And Shihab was cruelly denied his coffee...!

Outline

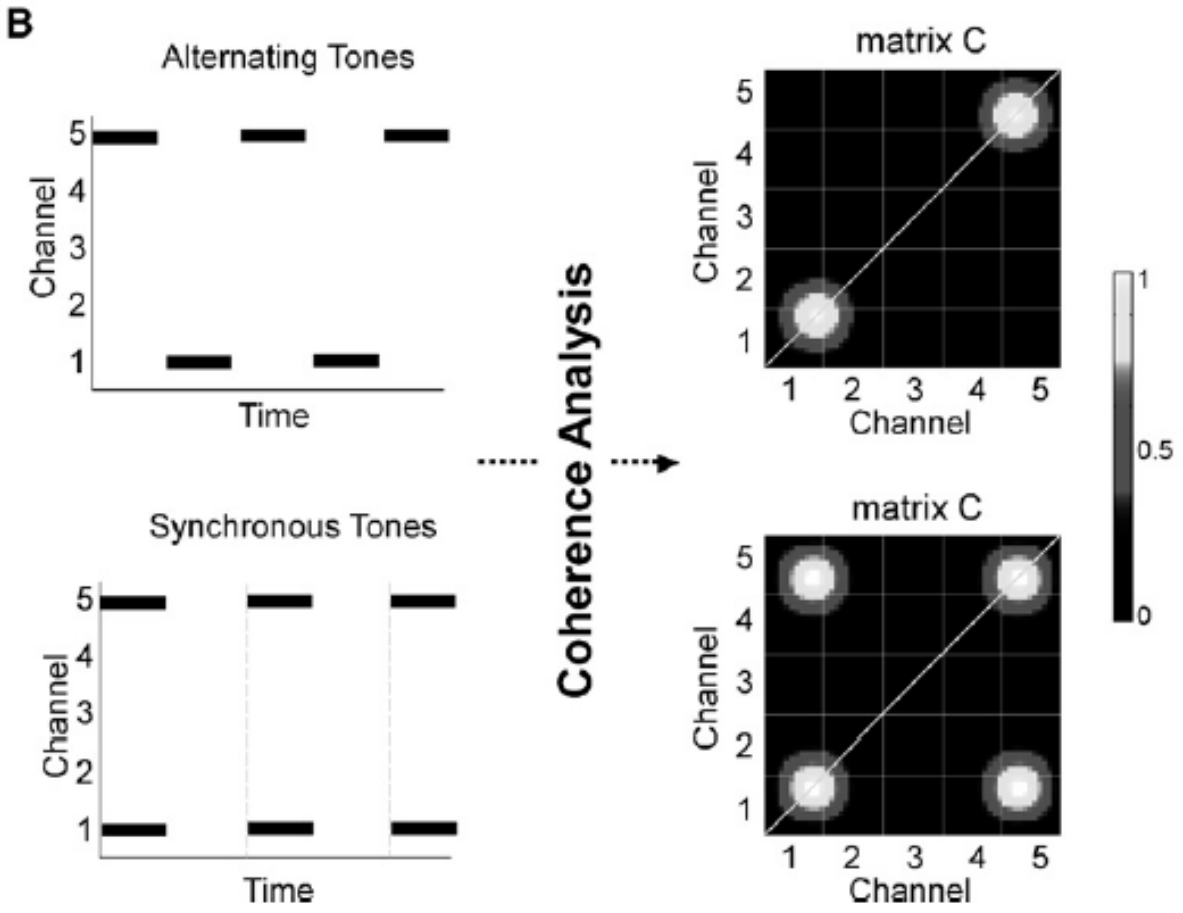
- Introduction
- Figure-ground stimulus
- fMRI study
- Psychophysics
- **Temporal coherence model**
- **Summary**

Temporal coherence model



(Chi et al., 2005; Elhilali et al., 2009; Shamma et al., 2011)

Temporal coherence model



(Elhilali et al., 2009)

Temporal coherence model

The temporal coherence model incorporates two different stages:

i. Feature analysis: -> multi-rate, multi-scale spectrotemporal receptive fields

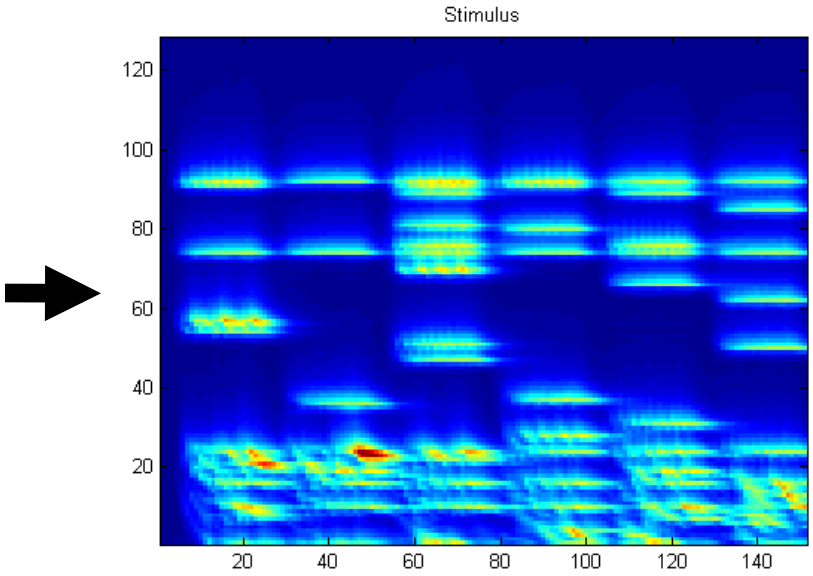
ii. Coherence analysis: -> dynamic coherence matrix

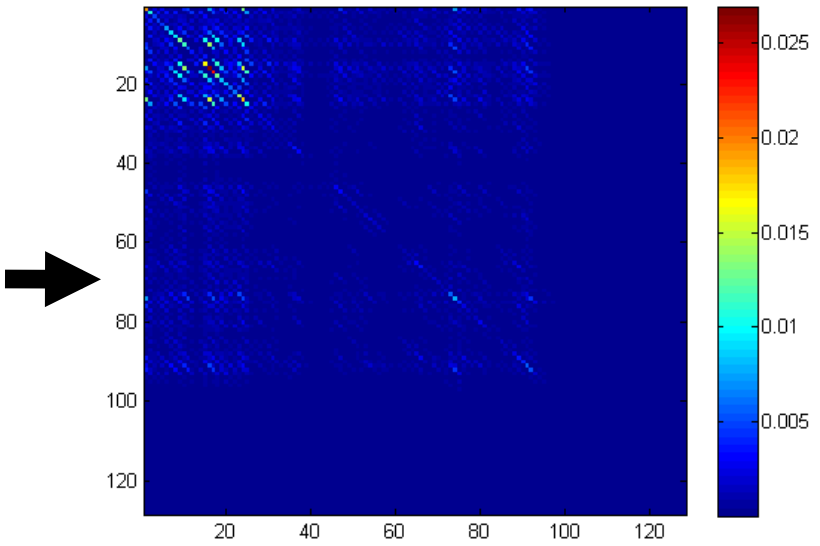
Analysis:

- **Parameters:** temporal modulation rate of 20 Hz, and bandwidth of 24 ch/octave
- **Input:** 1000 different examples of figure and ground stimuli for each (coh, dur)
- **Measure:** Average of maximum cross-correlation value for each stimulus
- **Output:** Average cross-correlation (figure) - Average cross-correlation (ground)

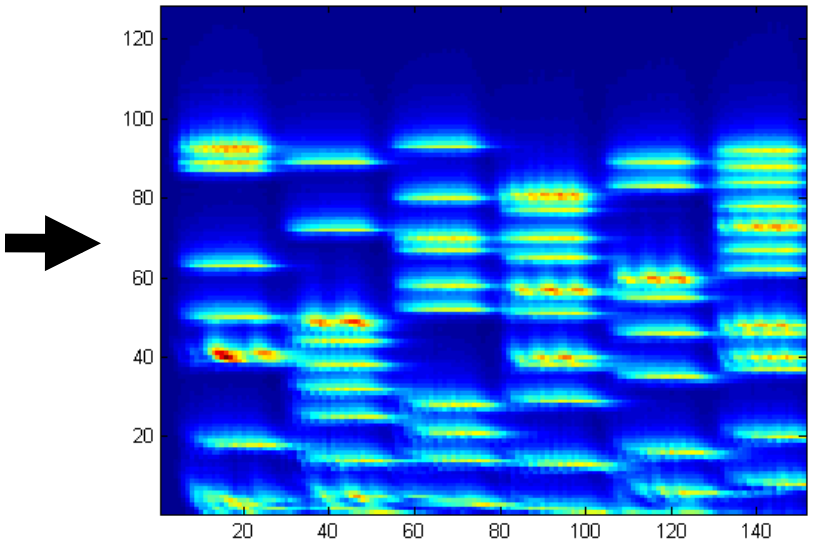
Isolated: expt. 5

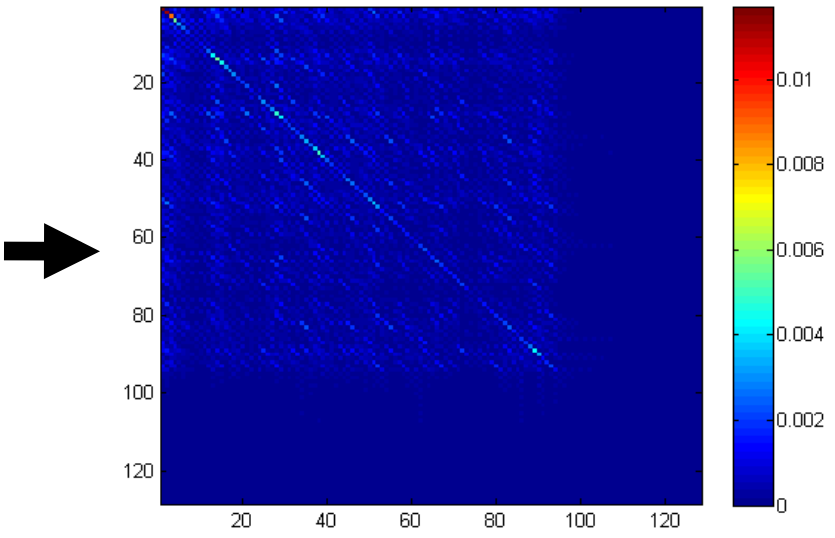
FIGURE





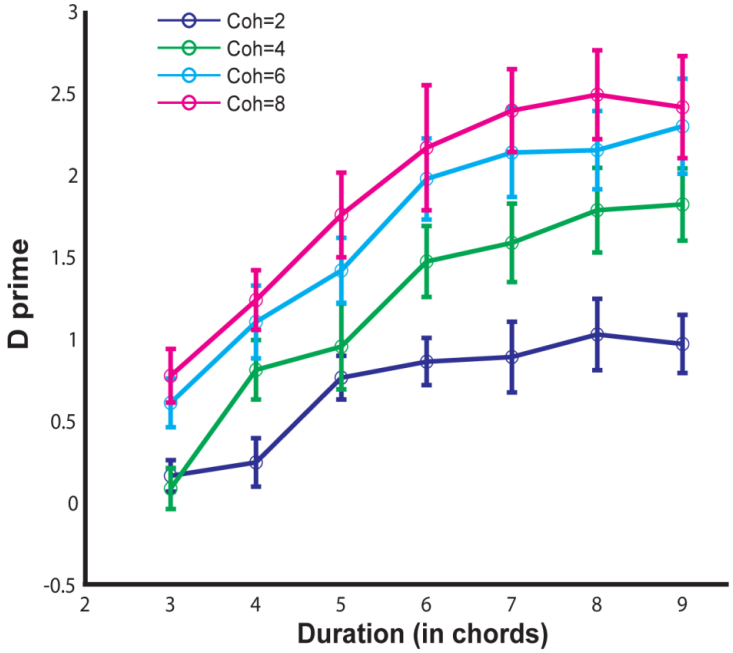
GROUND

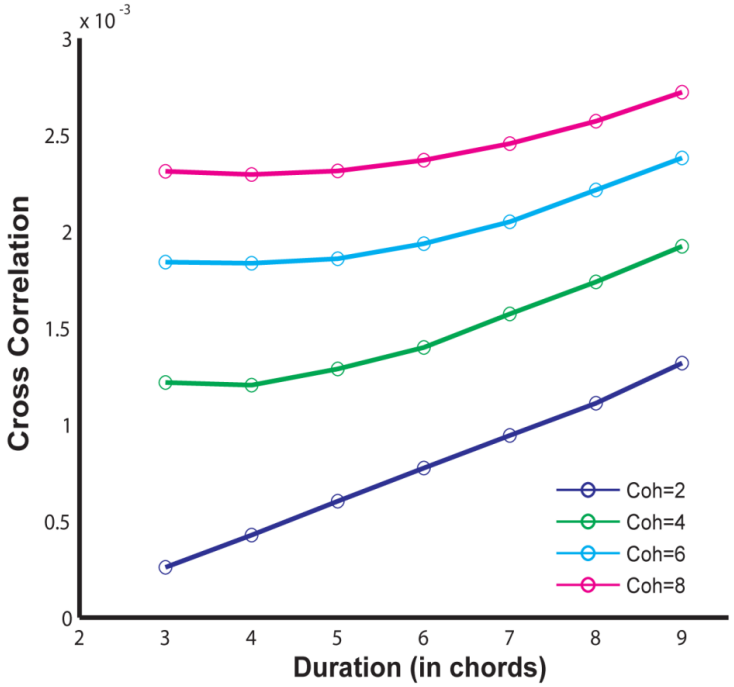




Expt. 5: Isolated

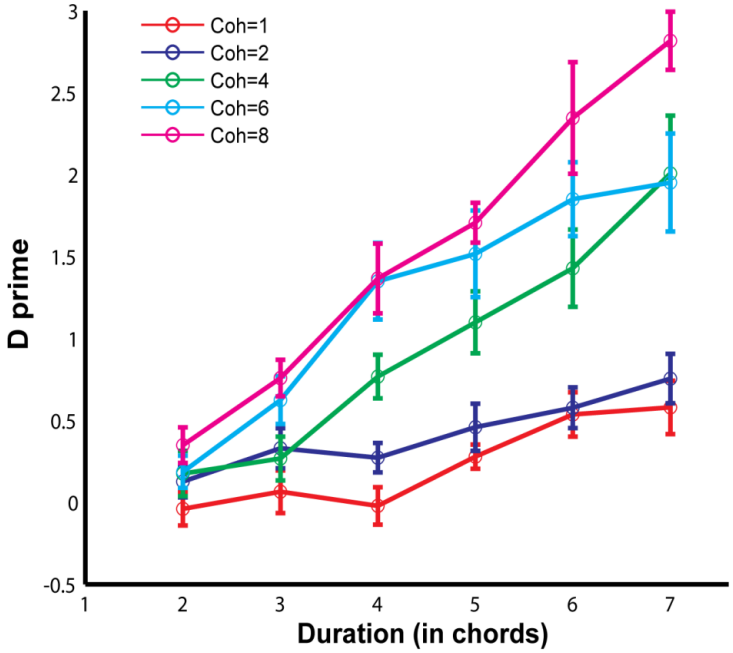
sv=24 Hz, rv=20 Hz

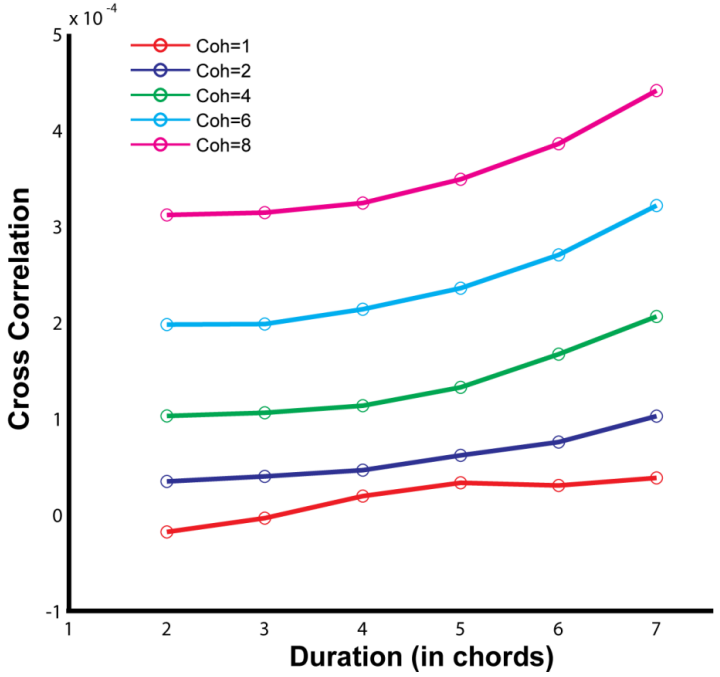




Expt. 1: Baseline (50 ms)

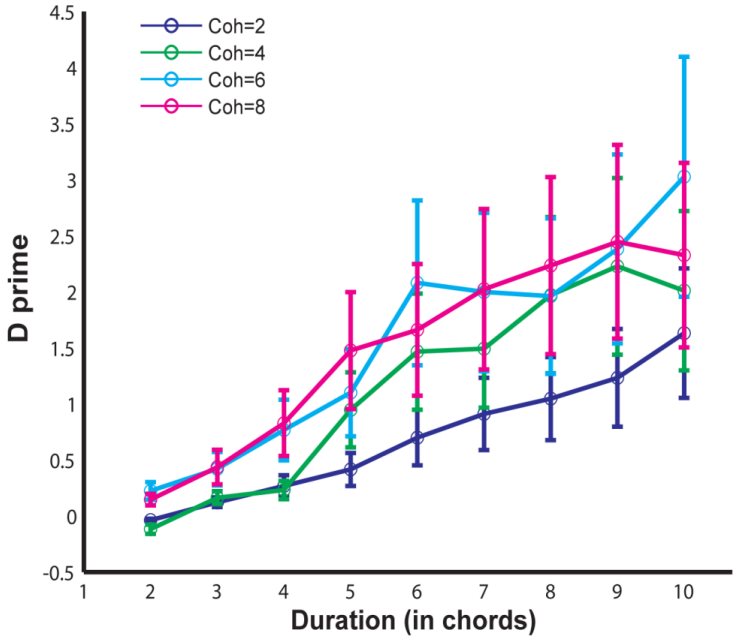
sv=24 Hz, rv=20 Hz

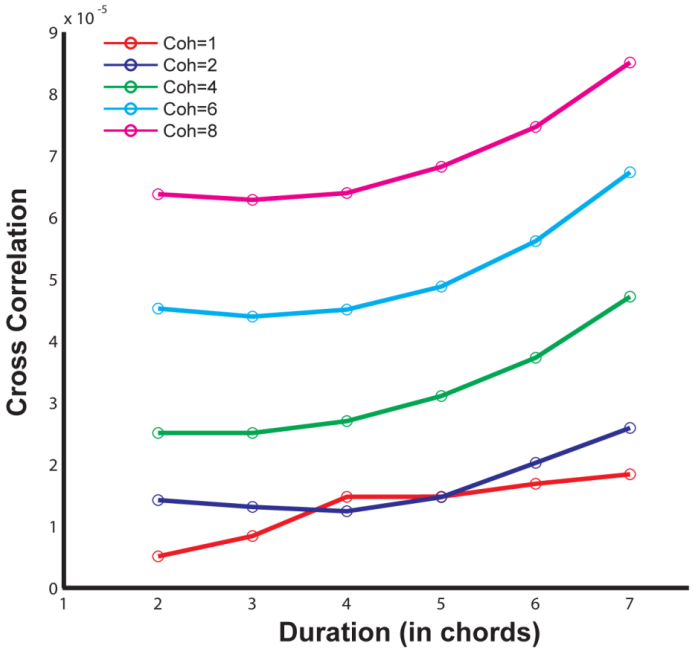




Expt. 2: Baseline (25 ms)

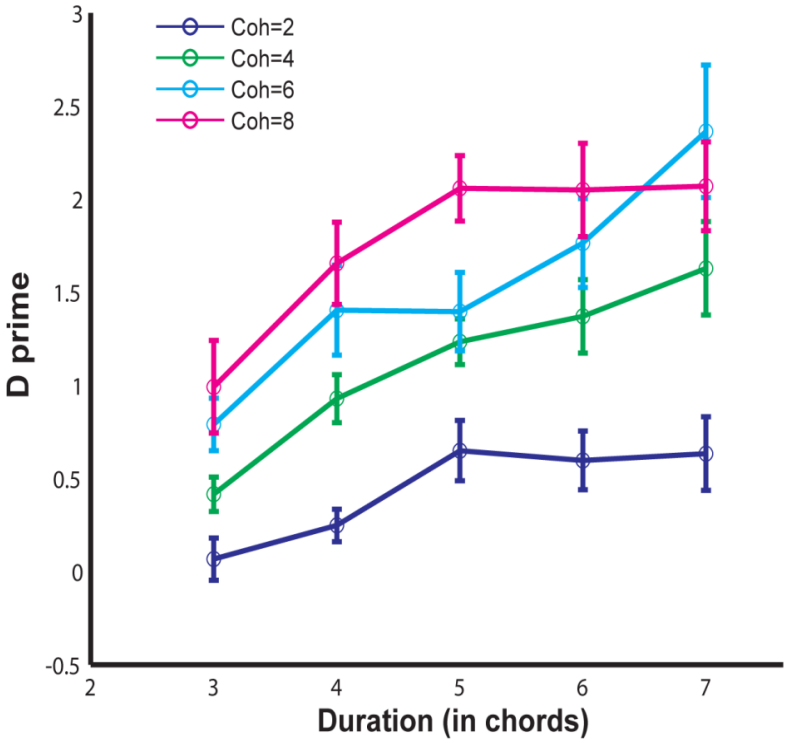
sv=24 Hz, rv=40 Hz

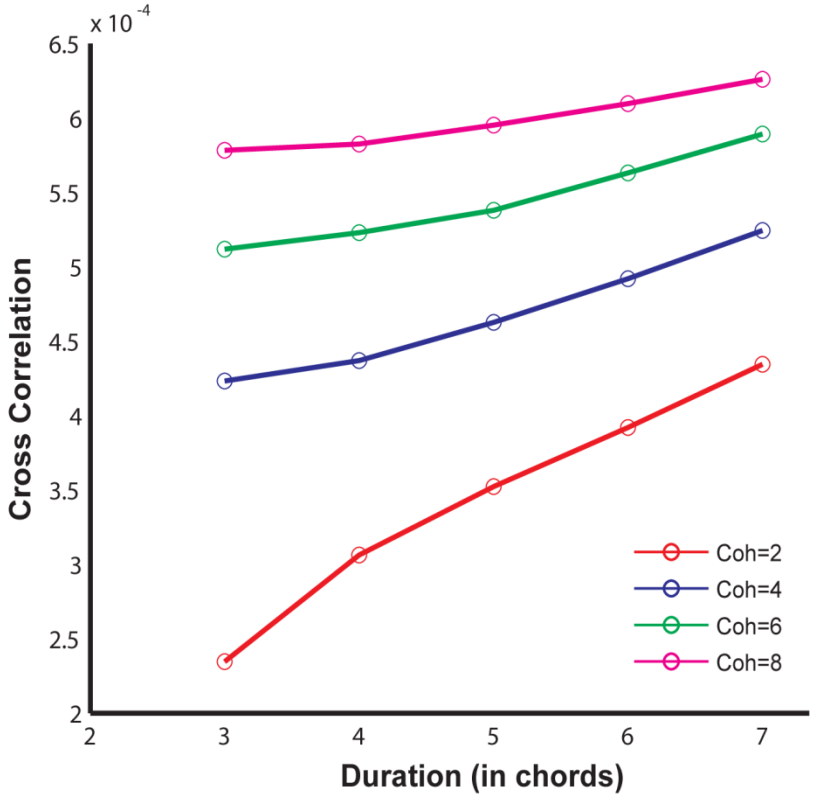




Expt. 3: SFG/Noise

sv=24 Hz, rv=20 Hz





Modelling summary

- Temporal coherence model can explain figure-detection in complex SFG stimulus for each psychophysics experiment.

- Model performs better than humans even at very short durations of the figure

- Auditory segregation in complex acoustic scenes may be based on computation of cross-channel coherence.

Outline

- Introduction
- Figure-ground stimulus
- fMRI study
- Psychophysics
- Temporal coherence model
- **Summary**

General summary

- **SFG stimulus:**
  Listeners can segregate figure from ongoing background very well.

- **fMRI:**
  Areas outside the auditory system, such as IPS are involved in segregation in complex acoustic scenes

- **Psychophysics:**
  Adaptation is not critical for complex auditory segregation.

- **Temporal coherence model:**
  Can explain figure-ground segregation in complex acoustic scenes.

Acknowledgments

Deborah Williams

Sukhbinder Kumar,

Timothy D. Griffiths

Newcastle Auditory Group

Aiysha Siddiq

Nicolas Barascud

UCL Ear Institute

Shihab Shamma

University of Marlyand, College Park

wellcome trust


WOMEN
KINGFISHER
KINGFISHER

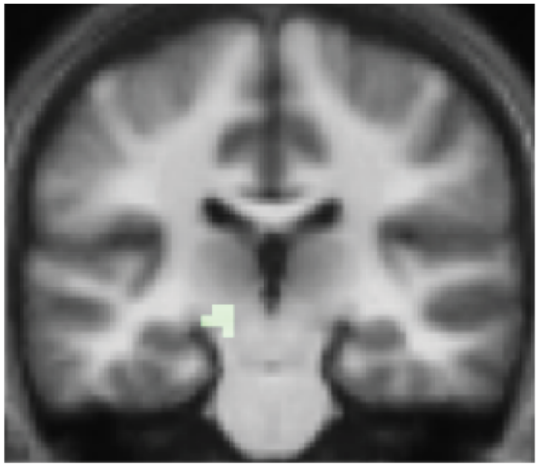
Questions ?

Table 1. Stereotactic MNI-coordinates

Contrast	Area	x	y	z	t	z
Effects of duration	Left IPS	-42	-46	64	5.14	3.67
		-48	-40	61	4.89	3.56
	Right IPS	51	-28	61	5.17	3.68
		45	-37	64	4.24	3.25
	Left STS	-57	-34	-2	4.42	3.34
	Right STS	60	-13	-11	4.06	3.16
	Right PT	60	-13	10	4.96	3.59
	Left MGB	-15	-25	-8	4.85	3.54
	Right MGB	18	-25	-8	4.92	3.57
Effects of coherence	Left IPS	-21	-73	46	4.99	3.60
		-24	-73	37	4.36	3.31
	Right IPS	27	-82	31	3.69	2.96
	Left STS	-48	-16	-5	3.43	2.81
	Right STS	39	-4	-26	3.77	3.00

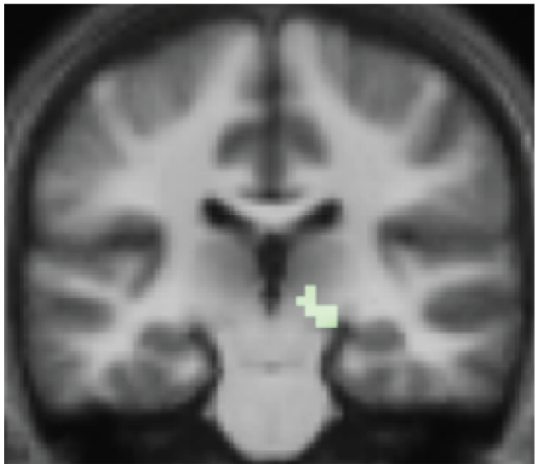
Local maxima for effects of duration, coherence as well as combined effects of duration and coherence are shown. Results are thresholded at $p<0.001$ (uncorrected)

Left MGB



y = -40

Right MGB



y = -37