Sundeep Teki [1]
Deborah Williams [2]
Nicolas Barascud [3]
Shihab Shamma [4]
Maria Chait [3]
Aiysha Siddiq [3]
Sukhbinder Kumar [1,2]
Tim Griffiths [1,2]

UCL

Temporal coherence mediates auditory object segregation in complex acoustic scenes

[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

Stimuli:

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

Mechanisms:

- frequency selectivity
- forward suppression
- neural adaptation

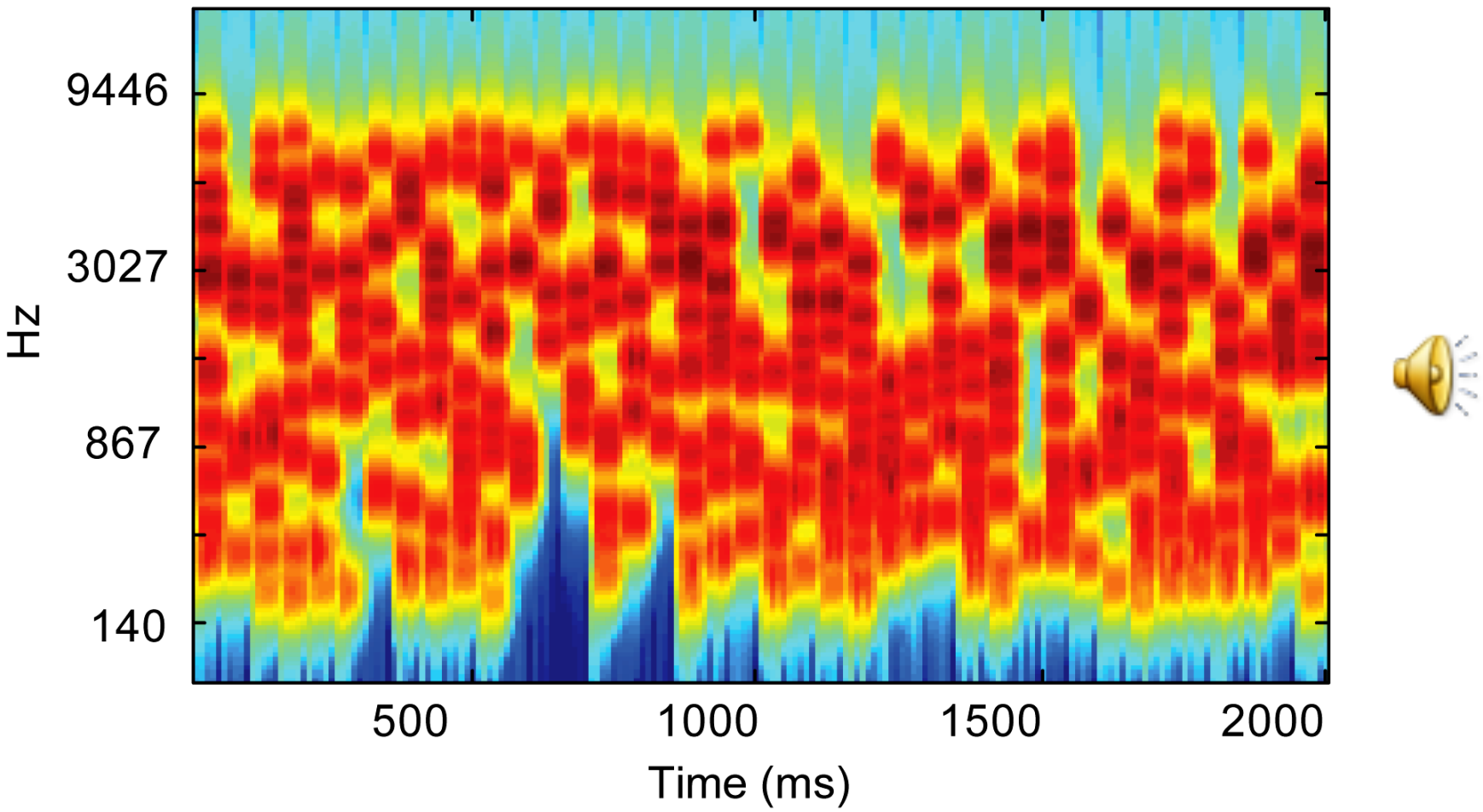
→ spatially segregated activation of neurons along the tonotopic axis corresponding to the two streams

Drawbacks of streaming signals:

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

Figure-ground stimulus

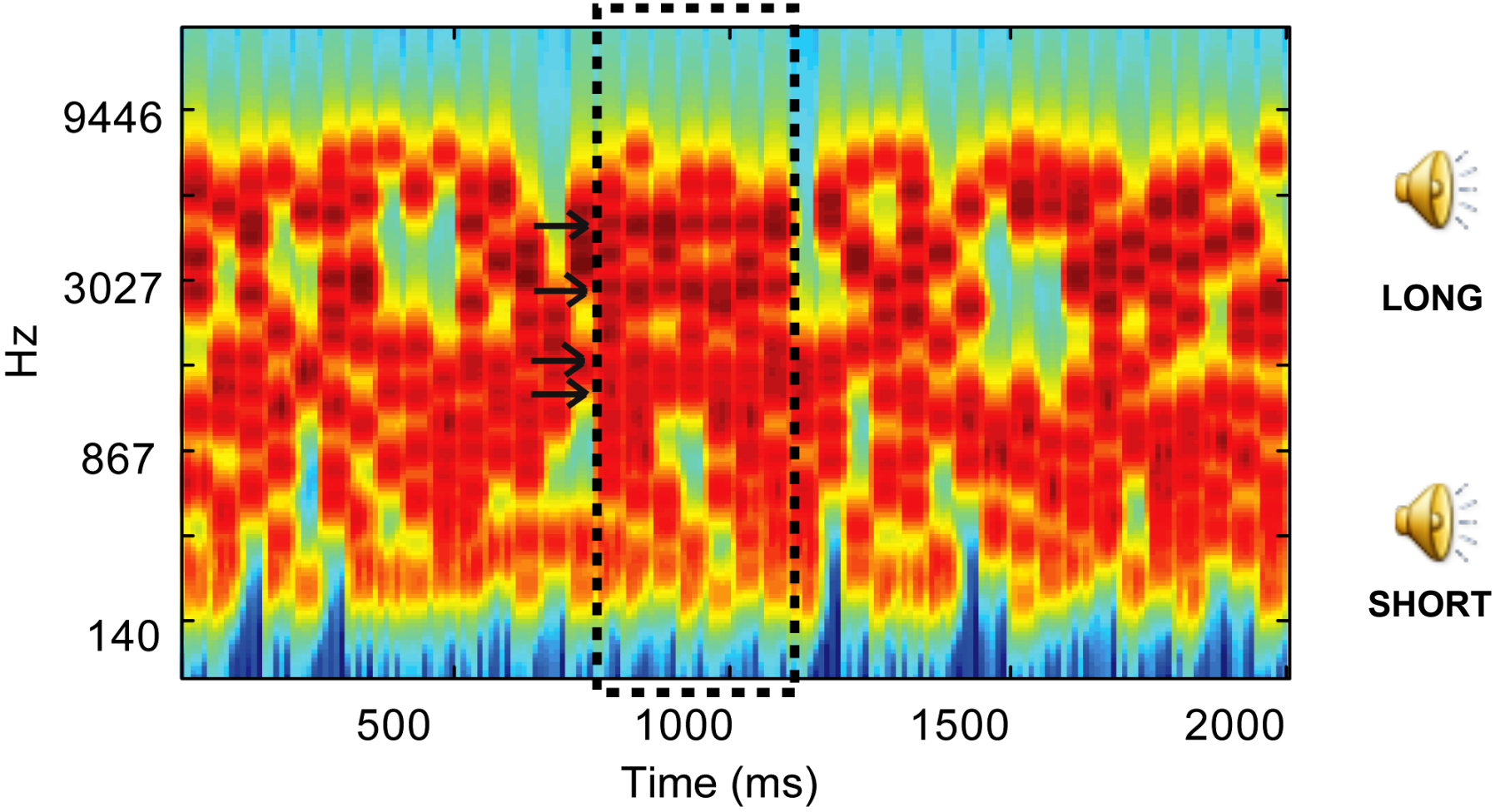
Stochastic Figure-Ground (SFG)



Teki, Chait et al., 2011. J Neurosci

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:	$1/24^{th}$ 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

Aims

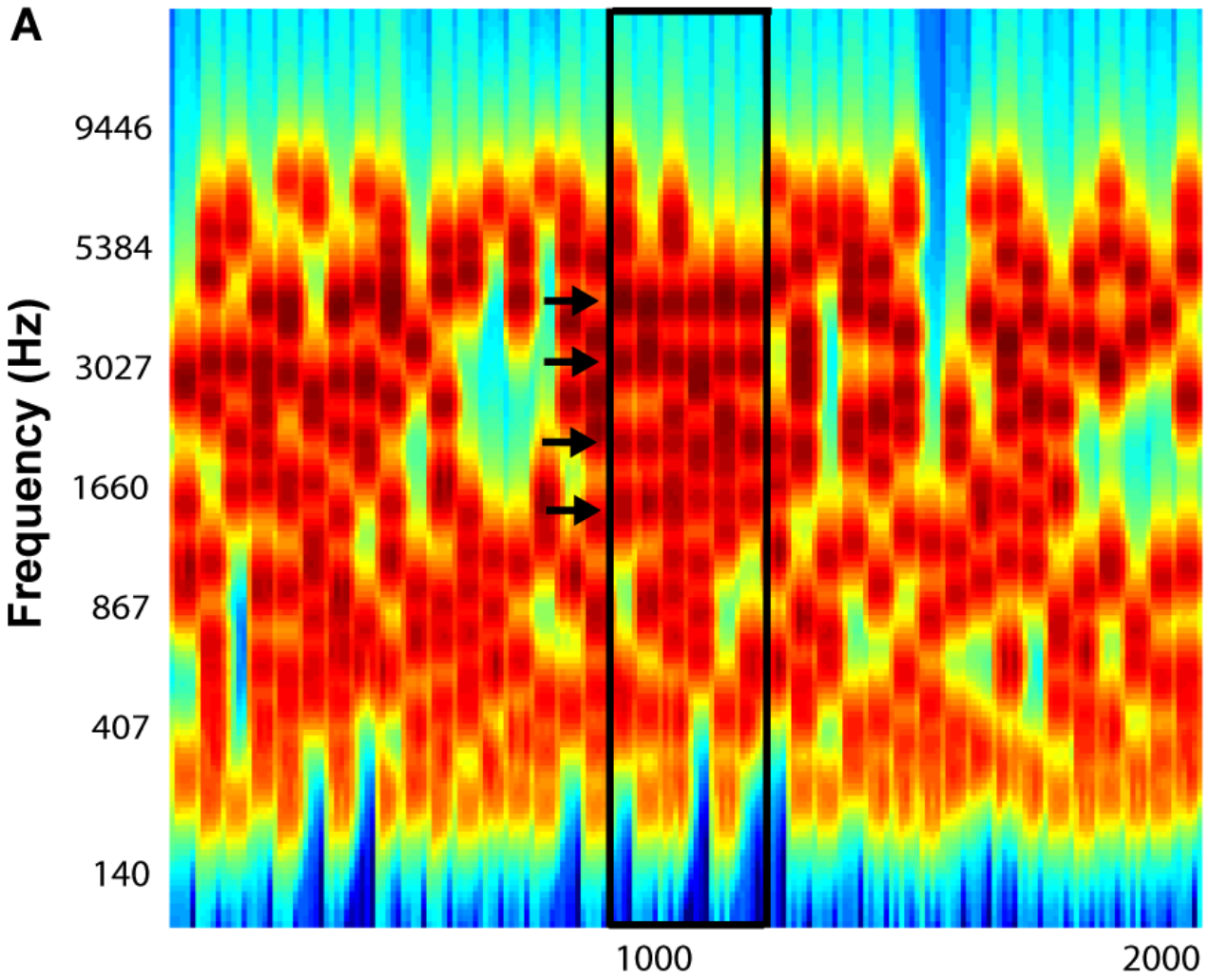
- To characterize the brain mechanisms that underlie segregation in complex acoustic scenes

- To examine listeners' ability to extract such complex patterns and test how robust their performance is to systematic stimulus manipulations

- Examine models of segregation to account for listeners' behaviour

Expt. 1: ‘Baseline’ (50 ms)

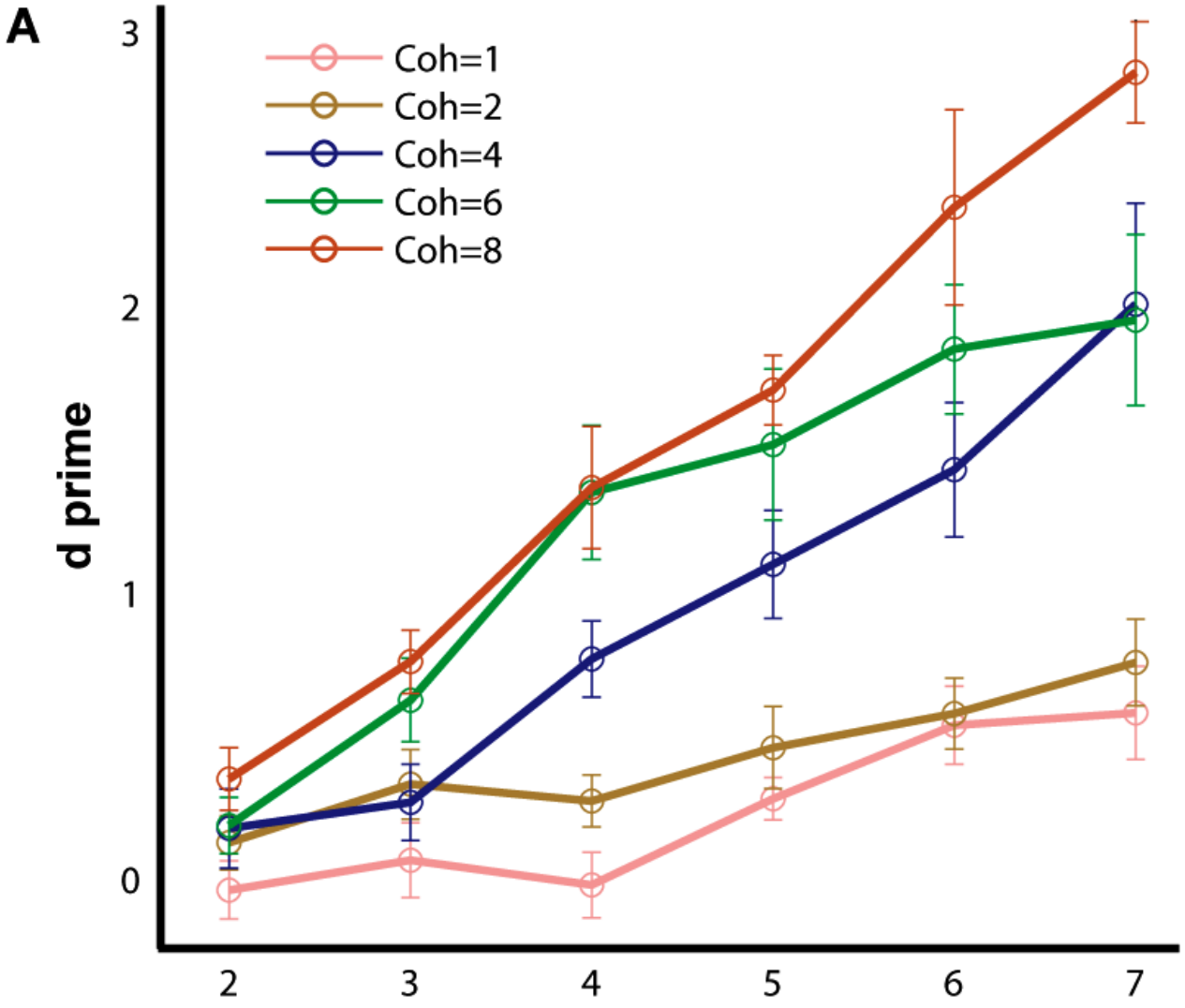
Stimulus consisted of a sequence of 40 x 50ms chords (2 s long)



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

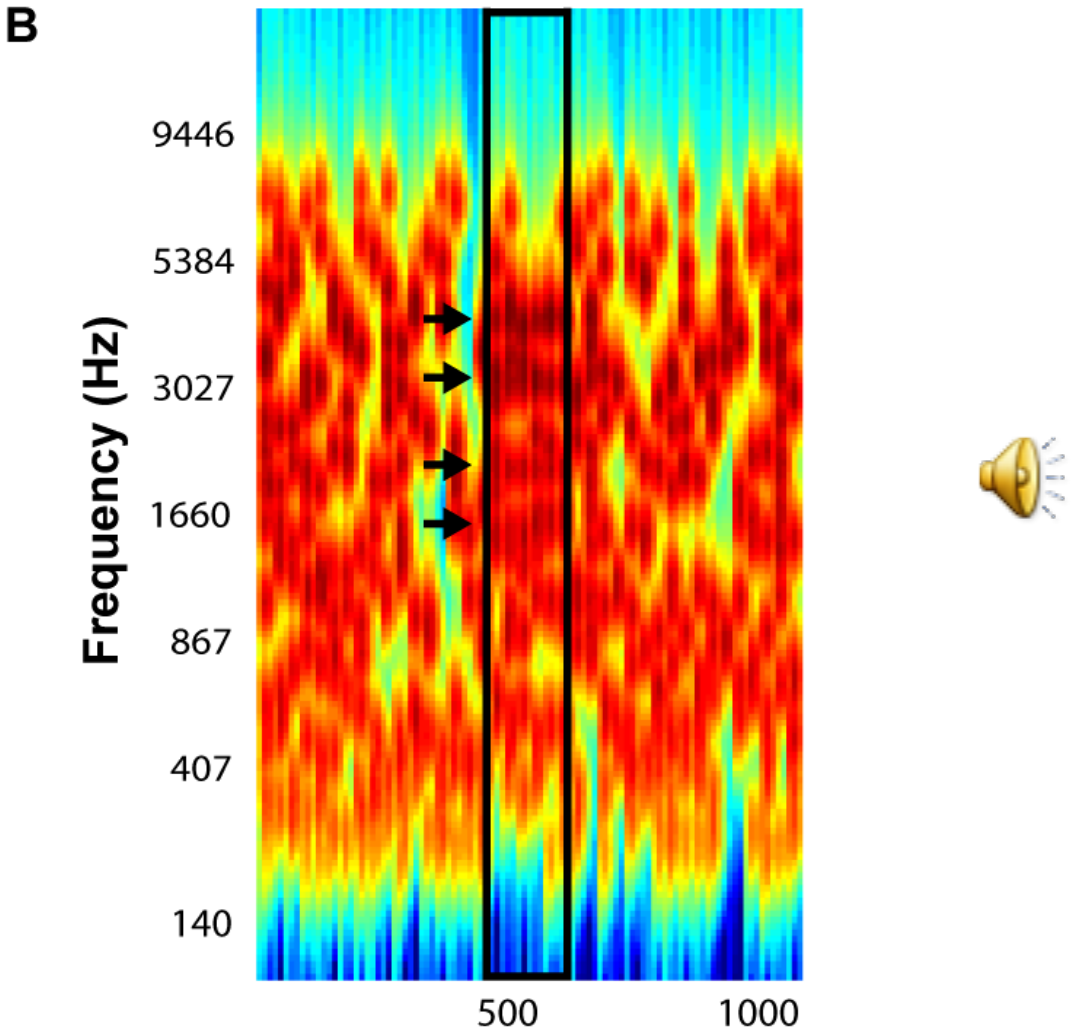
Expt. 1: Results

(n=9)



Expt. 2: ‘Baseline’ (25 ms)

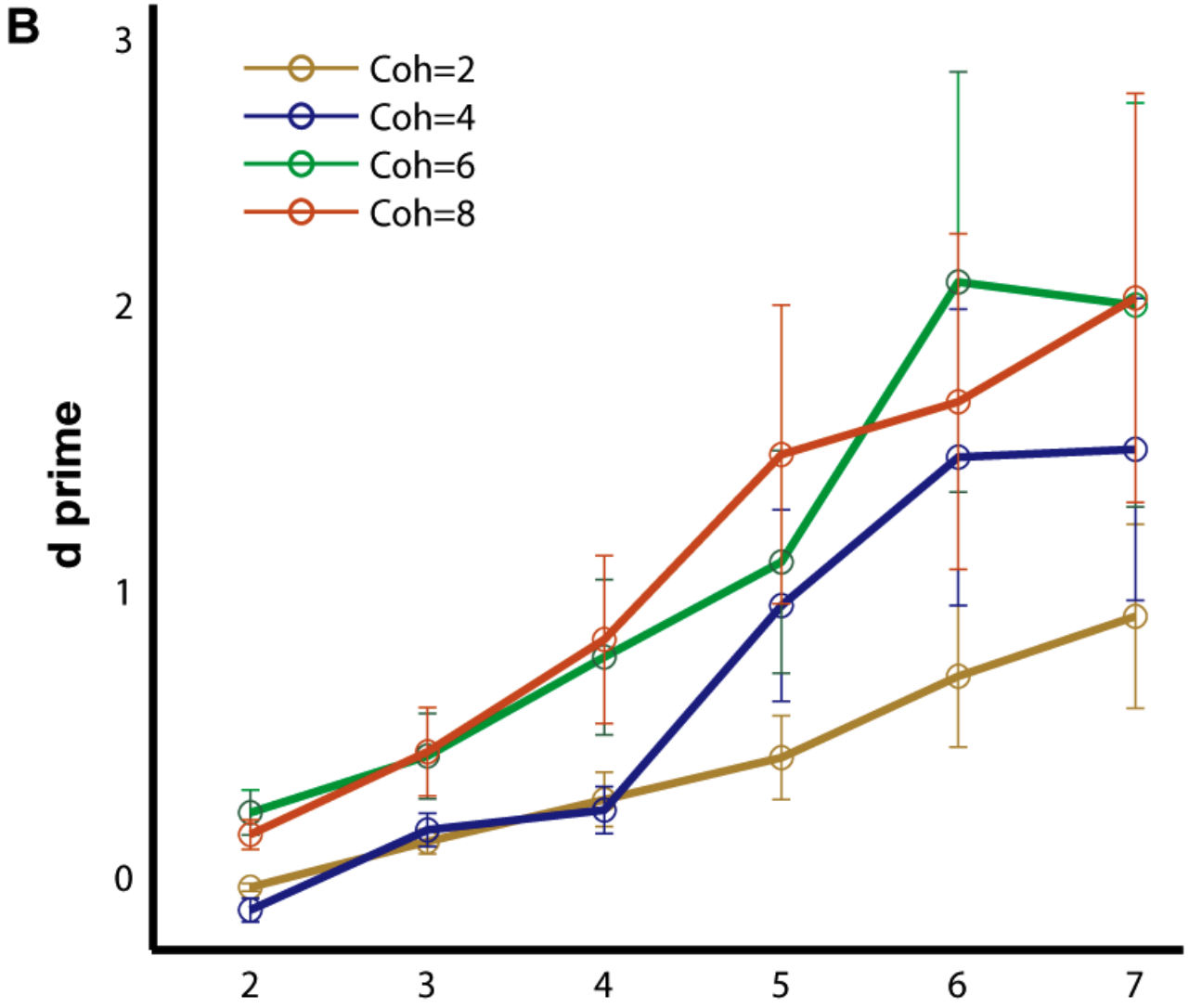
Stimulus consisted of a sequence of 40 x 25ms chords (1 s long)



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

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

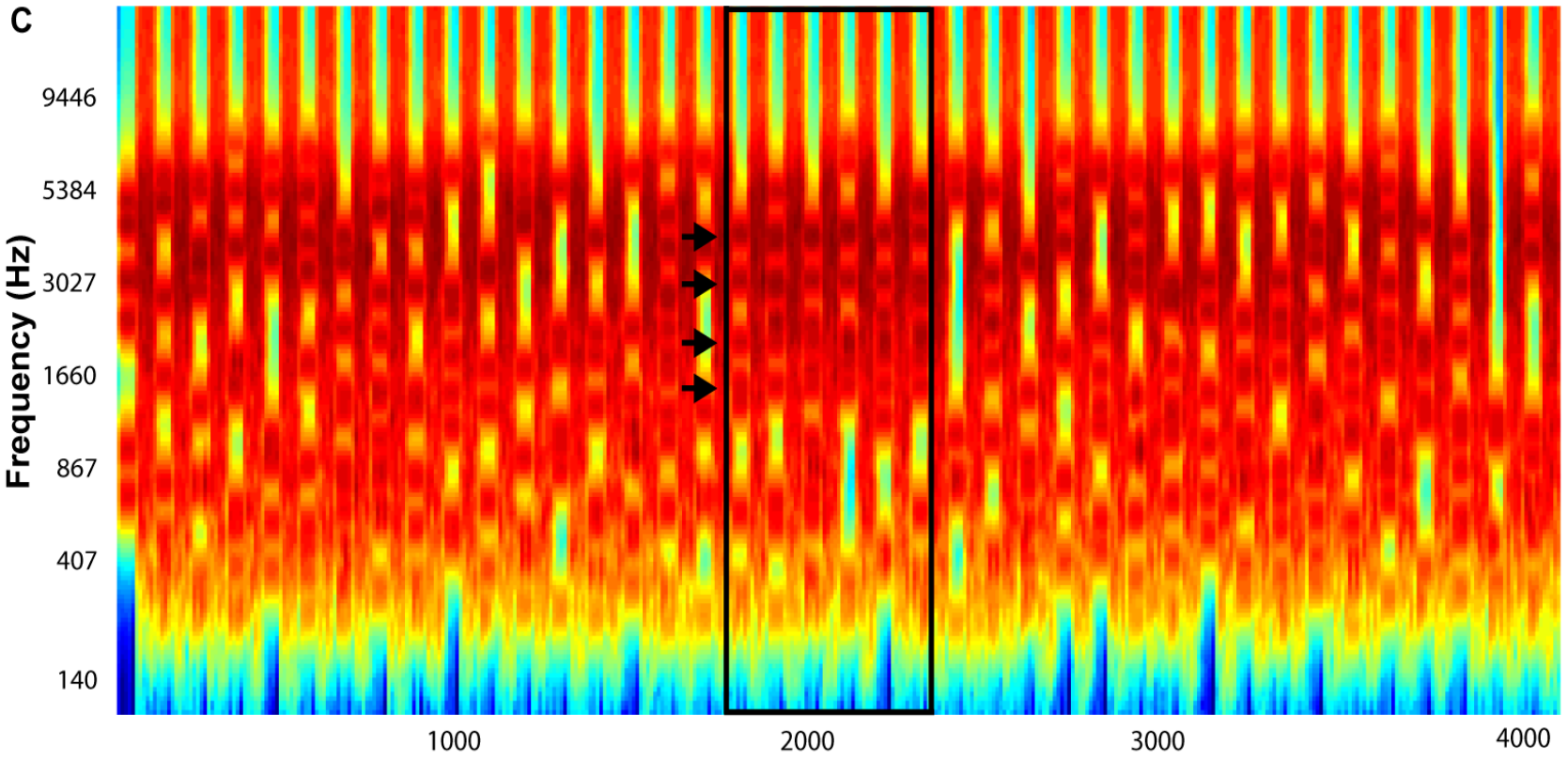
Significant effect of coherence: $F(3, 45) = 77, p < 0.001$

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’

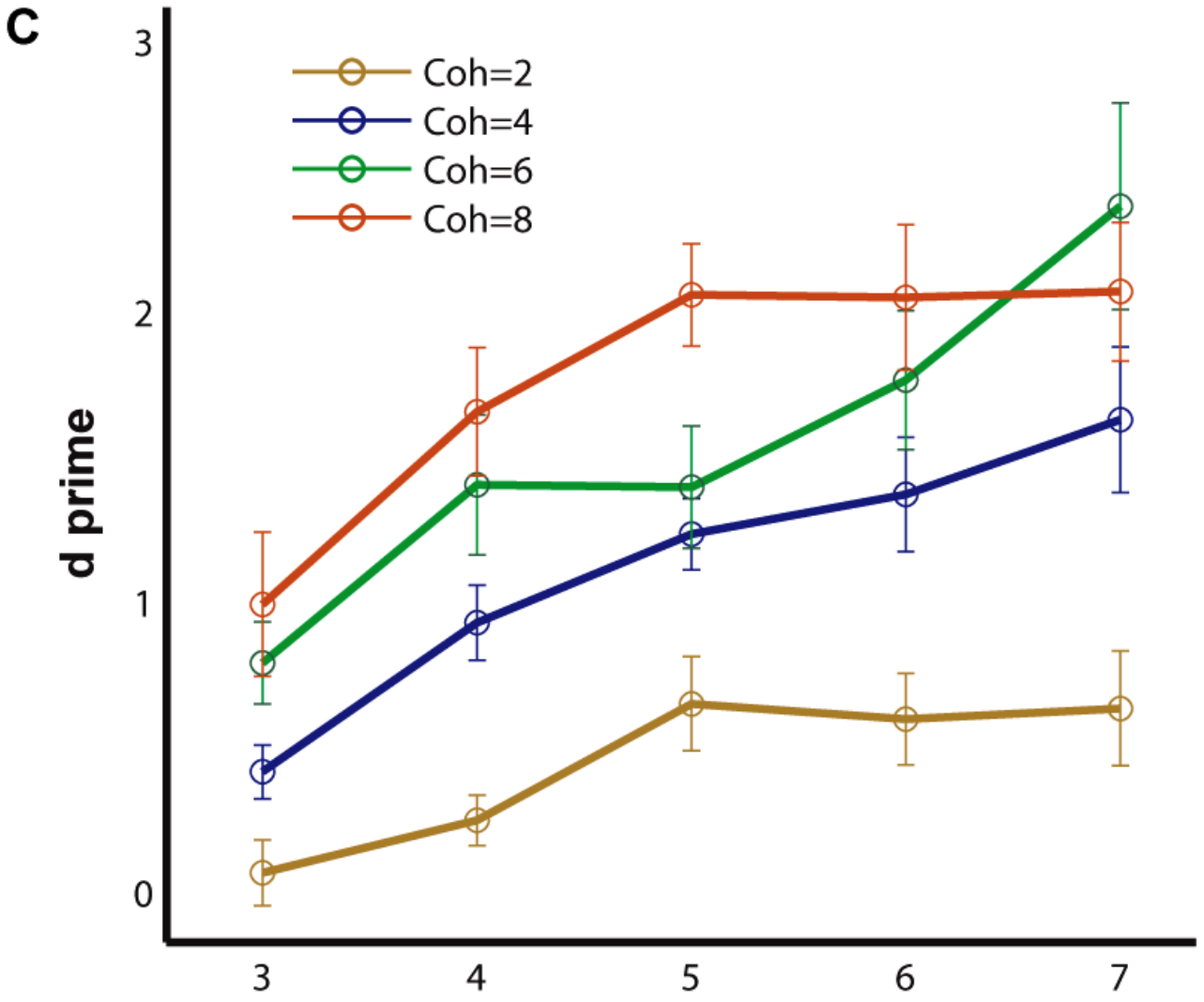
Stimulus: SFG with 40 x 50ms chords alternating with 50ms of white noise (4 s)



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

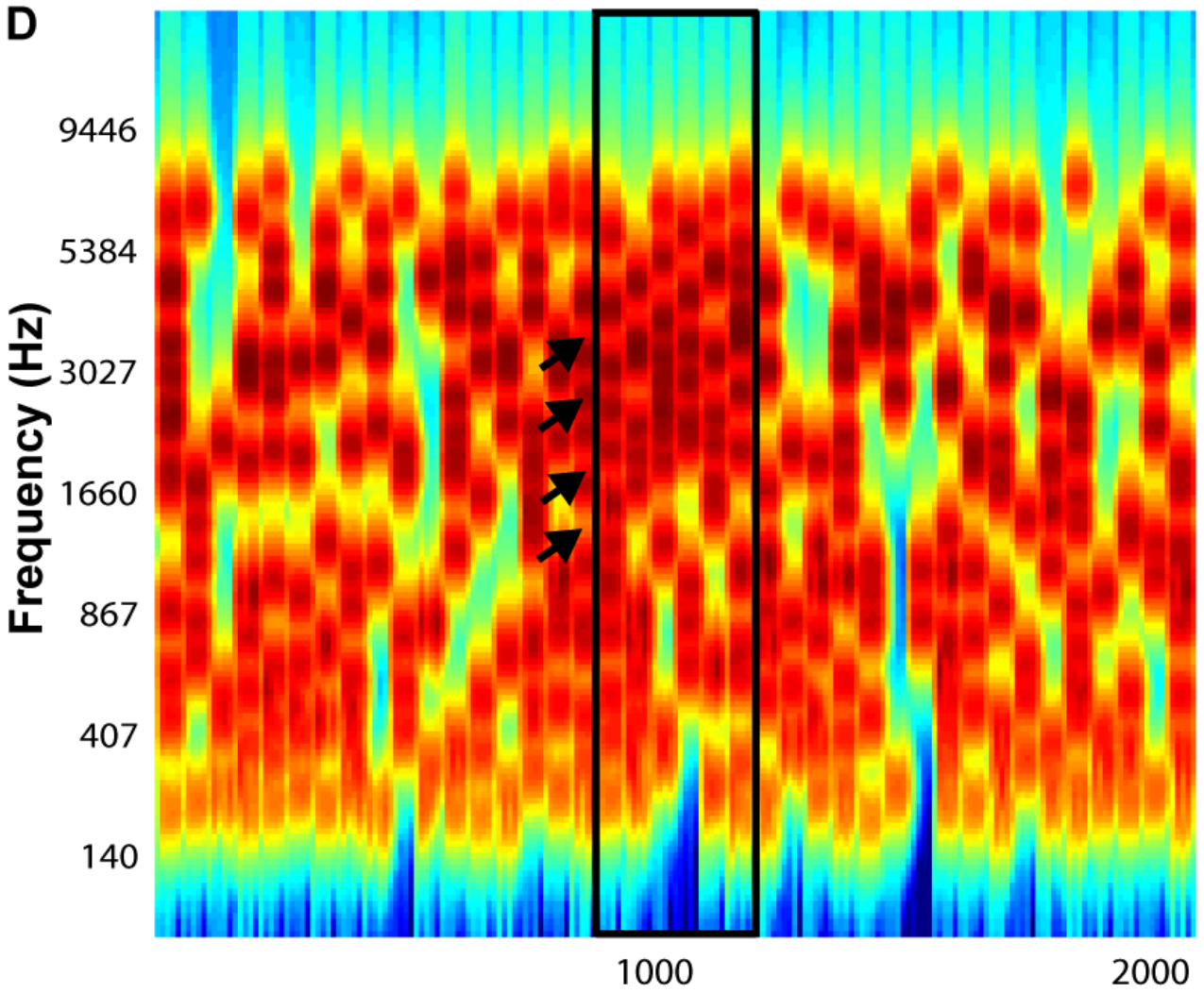
Results

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: $F(1,17) = 0.004, p = 0.953$

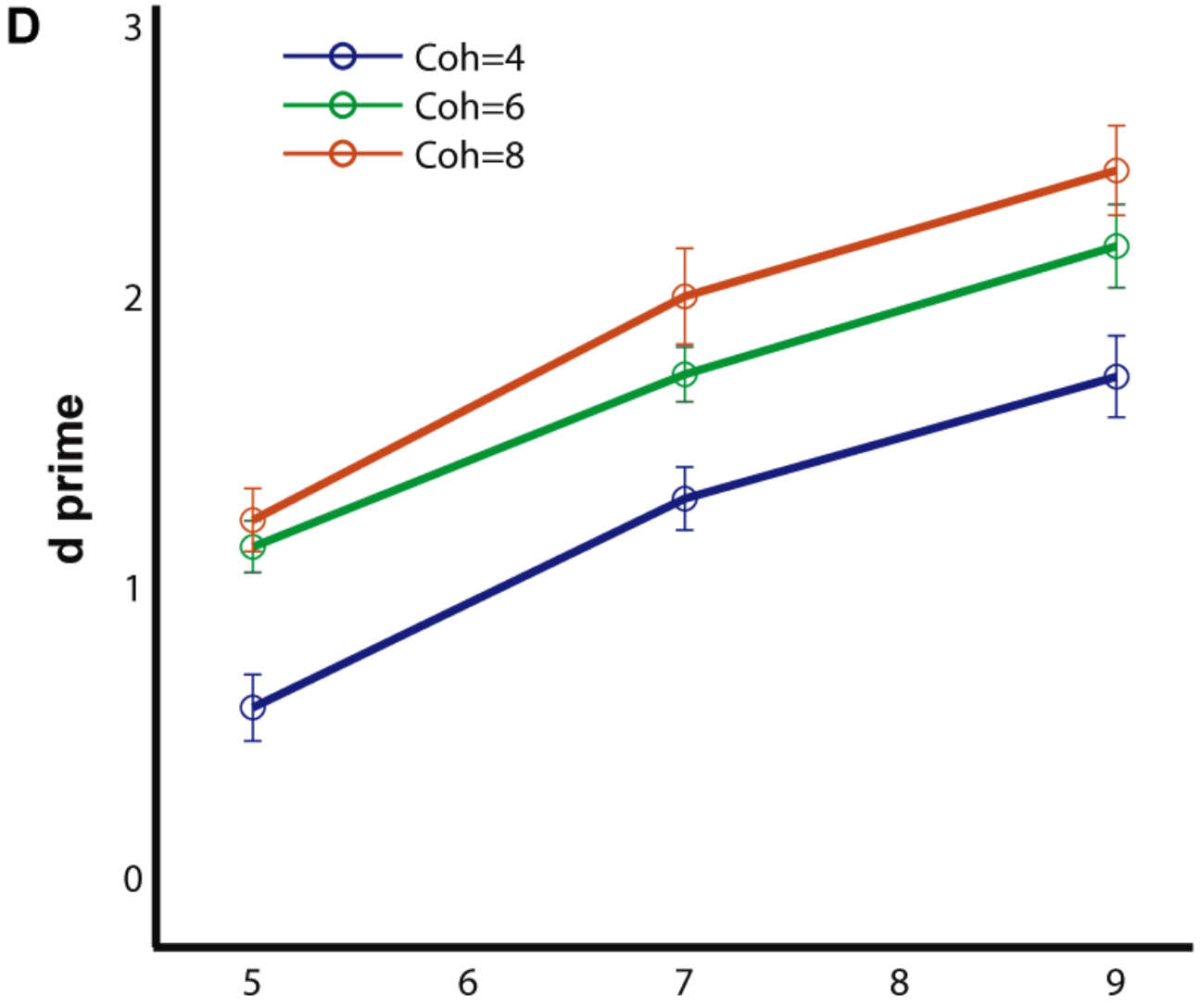
Expt. 4: ‘Ramps’



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

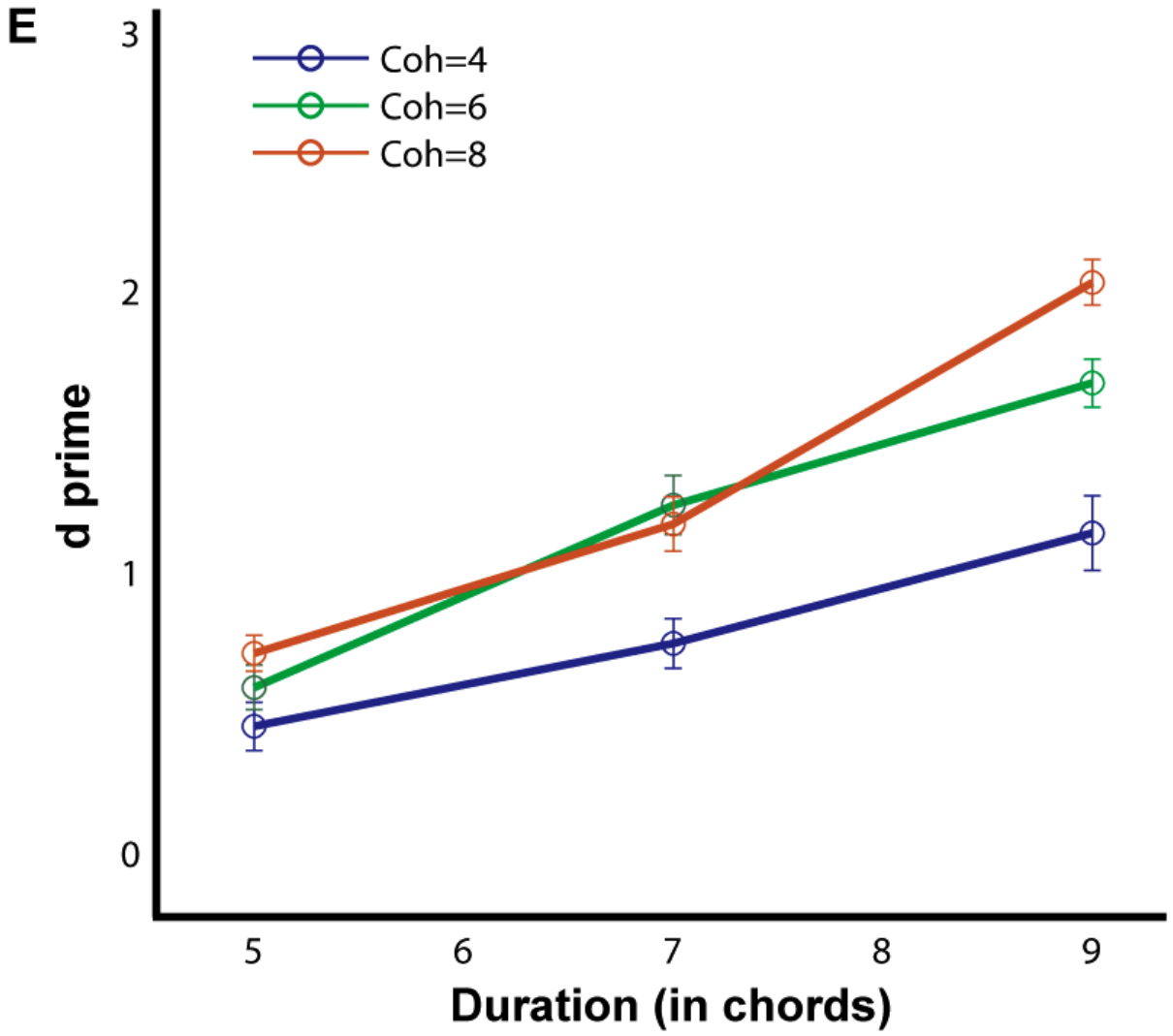
Results: Ramps 2

(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. 5: ‘Isolated’

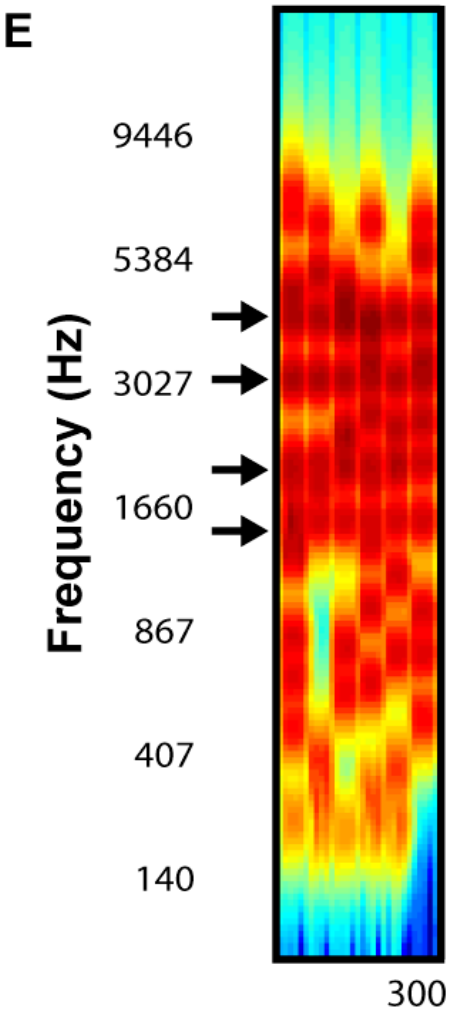


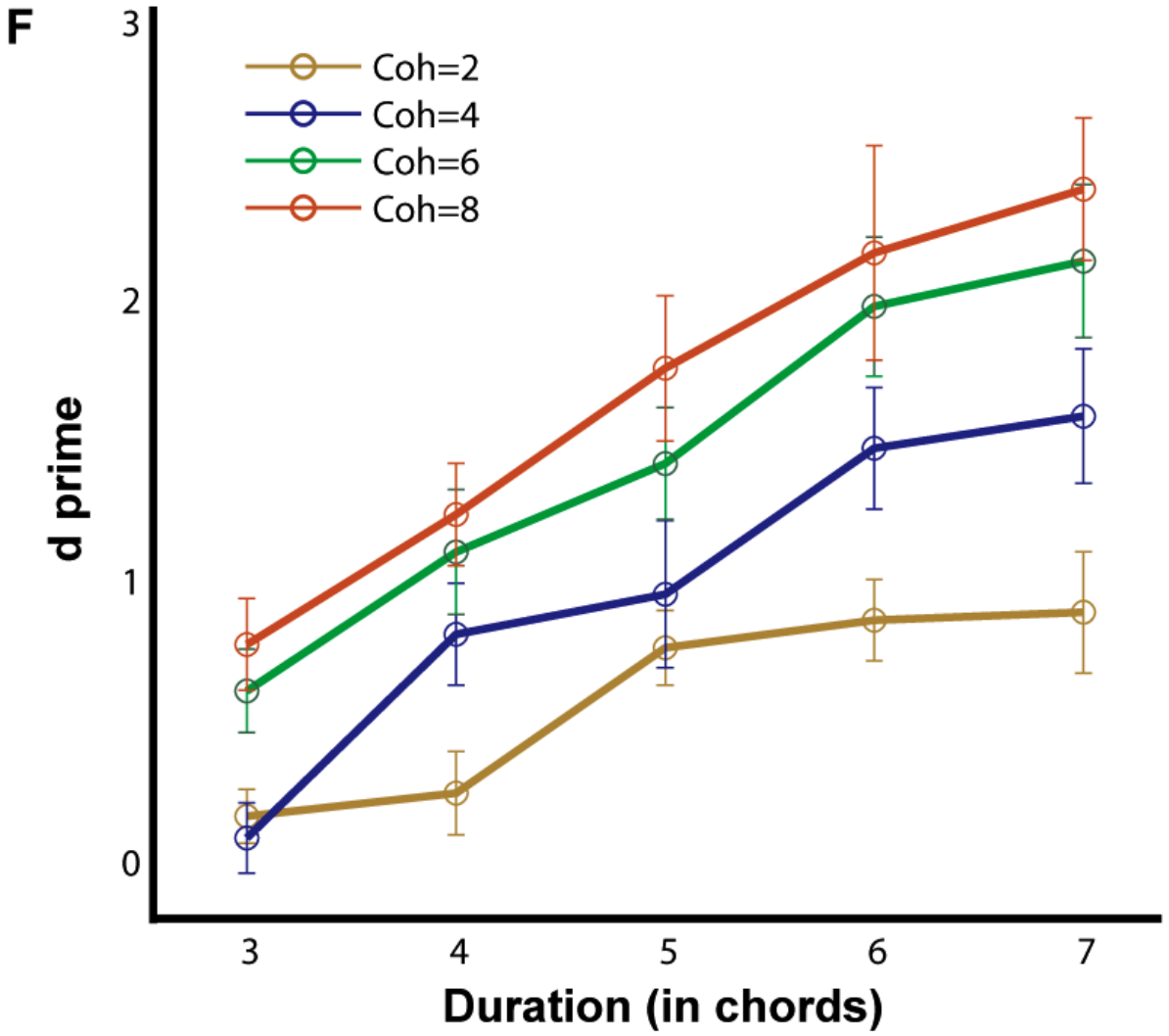
Figure:

Ground:

Coherence: [2 4 6 8] **Duration: [3:7]**

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

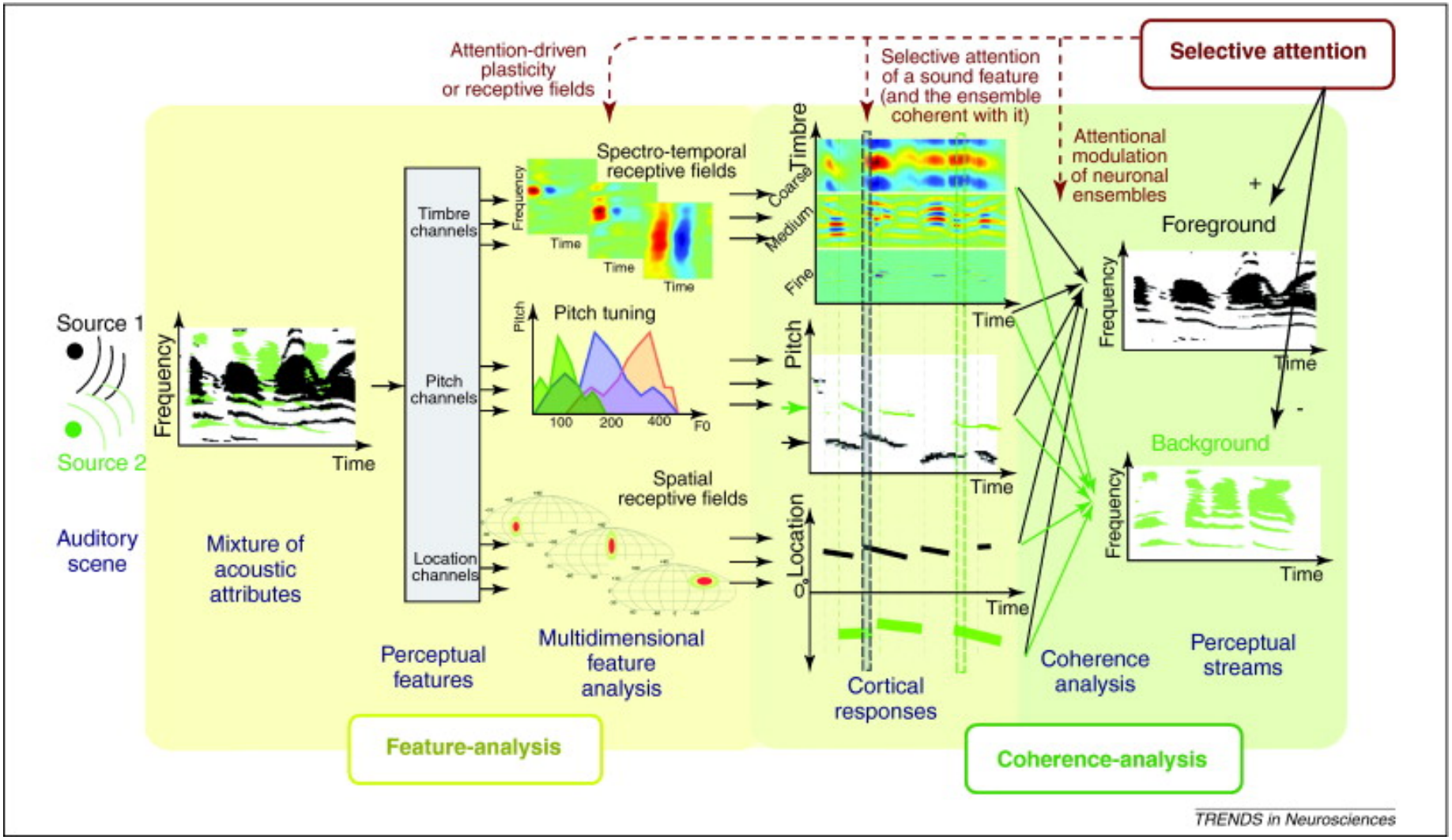
Psychophysics summary

Figure-detection performance in complex SFG stimulus is:

- **Dependent 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)
- **Invariant to the presence of preceding background** (Expt. 1 & 5)

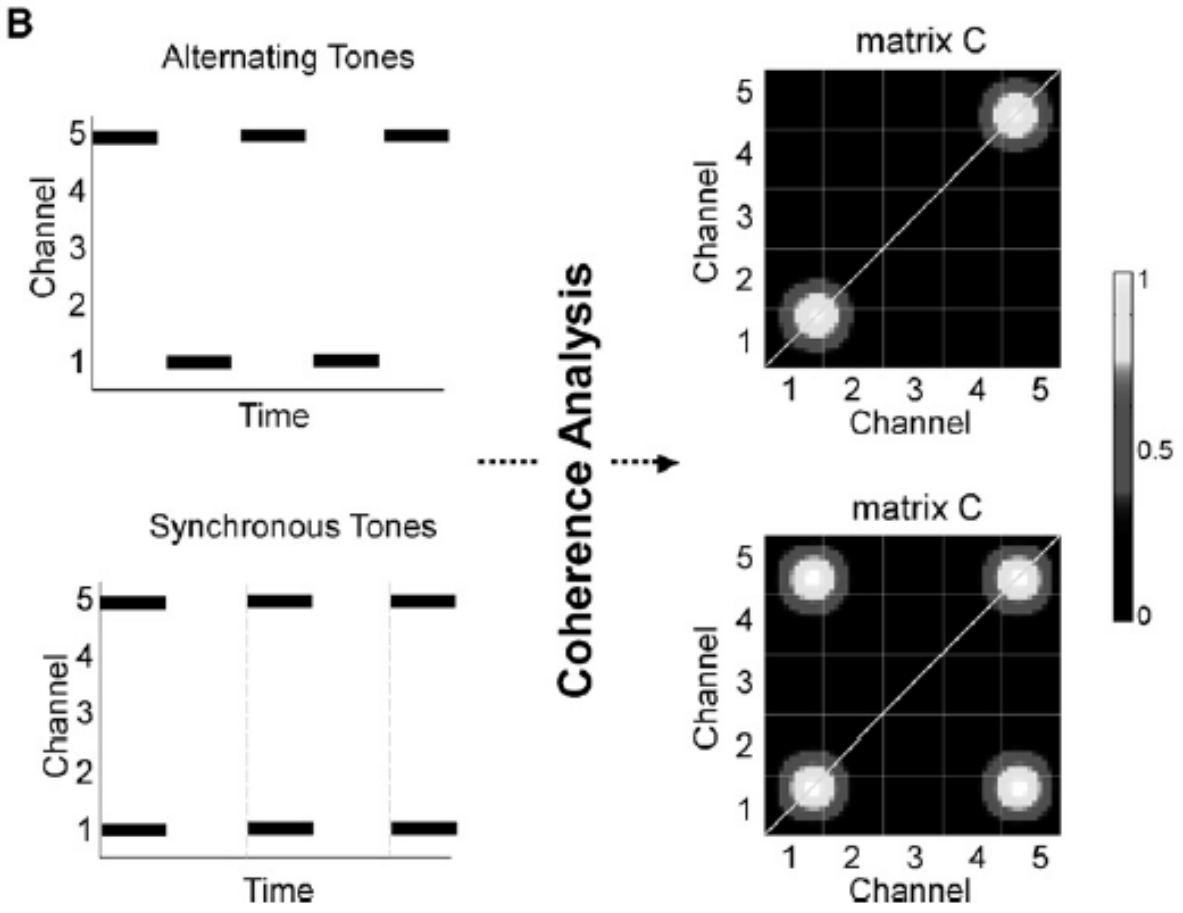
Temporal coherence modelling

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

Hypotheses:

Channels with repeating frequency components would be temporally coherent; and these components may be grouped together and perceived as a single object.

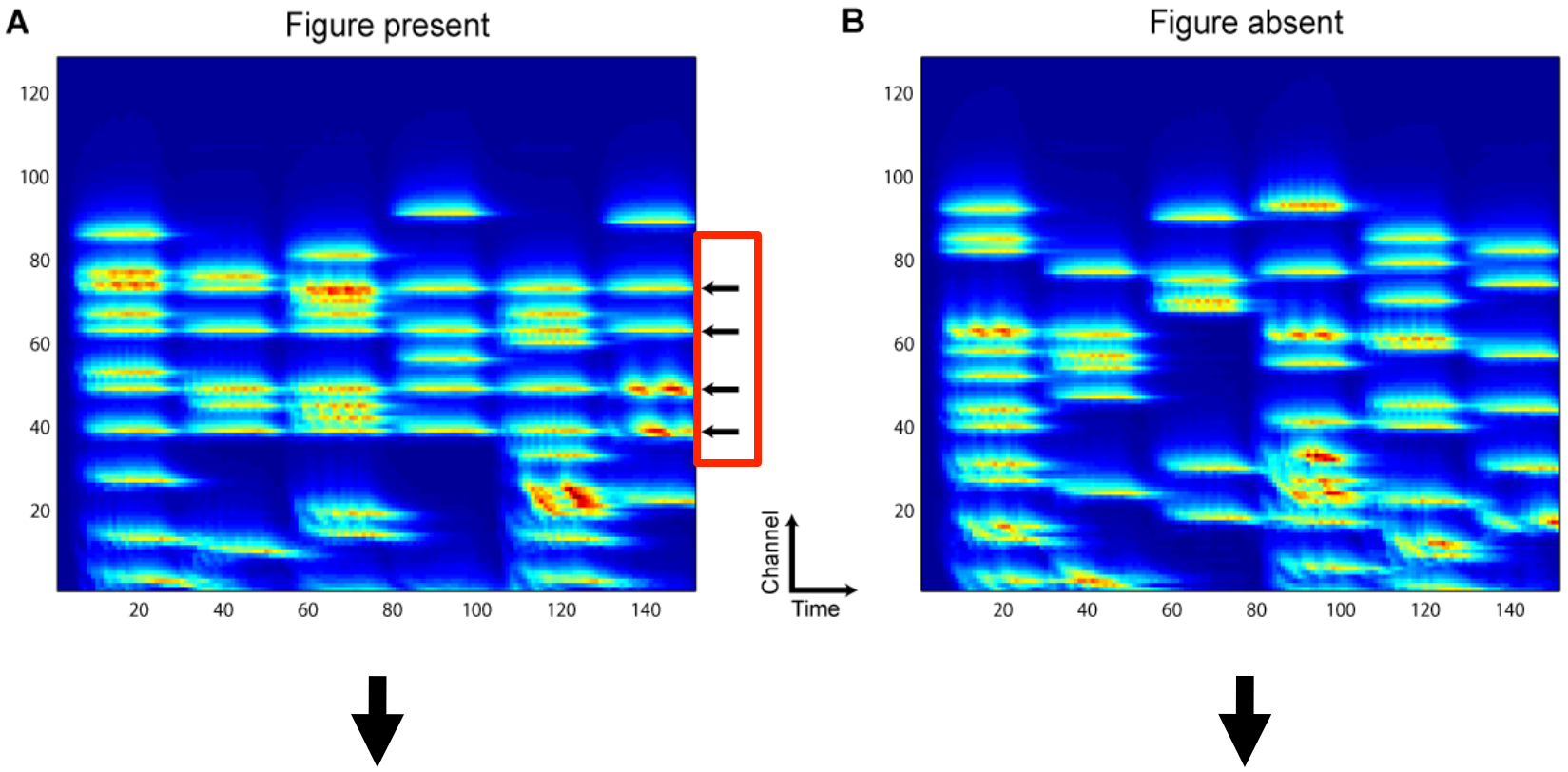
Parameters of the model:

Temporal modulation rate: **20 Hz** (tuned to chord repetition rate of 50 ms)

Spectral resolution scale: **8 cyc/oct.** (corresponding to BW in streaming)

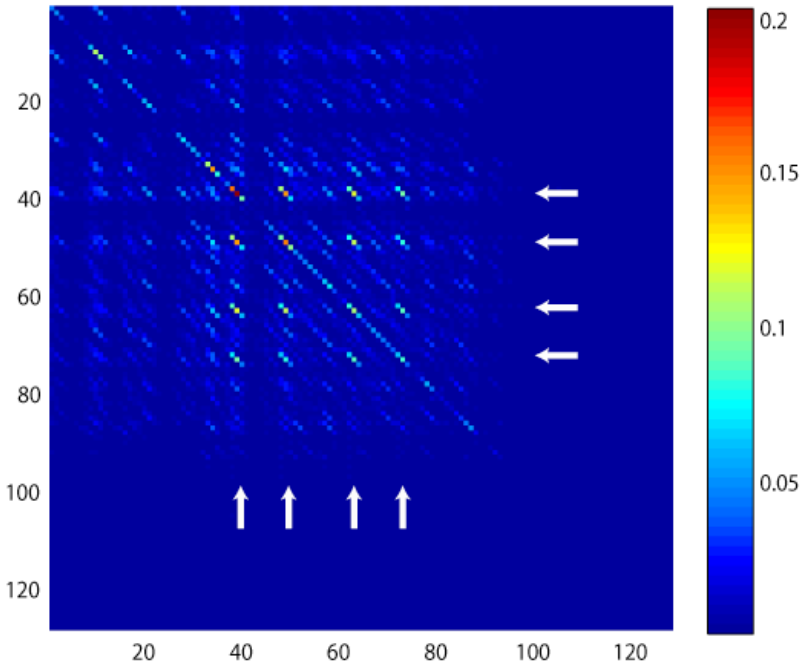
Isolated: expt. 5

I. Input: **1000 different examples of figure and ground stimuli for each (coh, dur)**

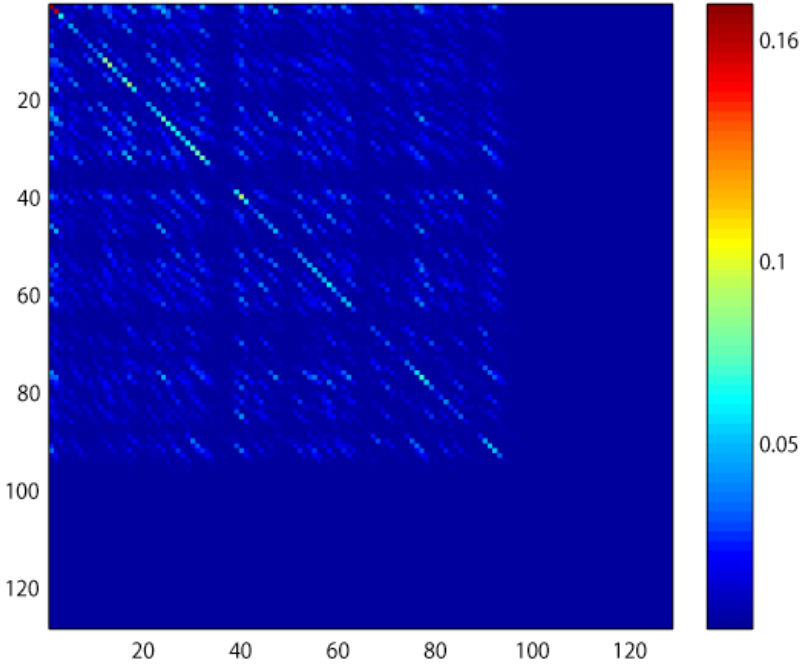


Isolated: expt. 5

Coherence matrix
(figure present)



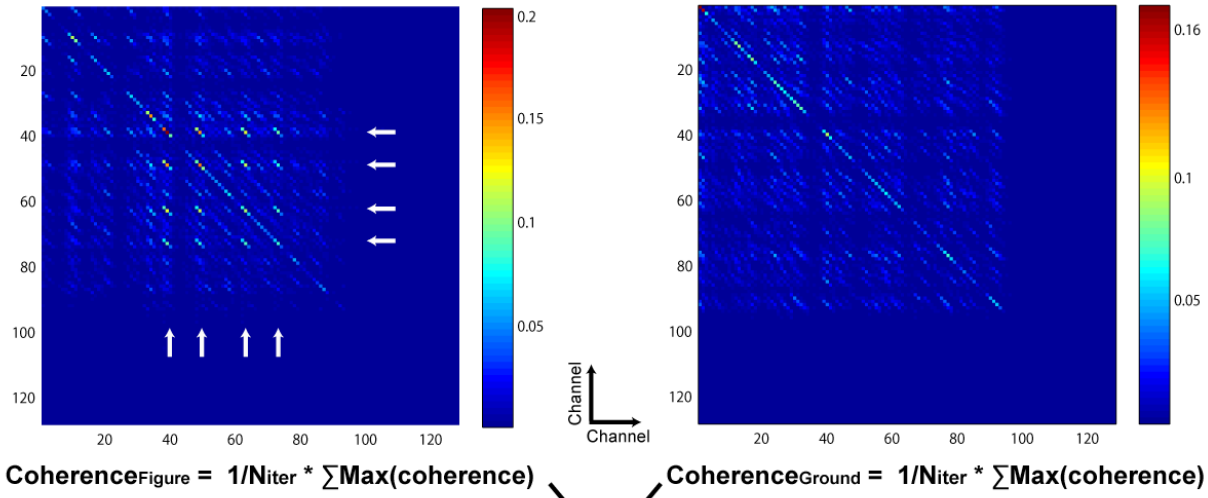
Coherence matrix
(figure absent)



- **Measure:** **Maximum cross-correlation value for each stimulus**

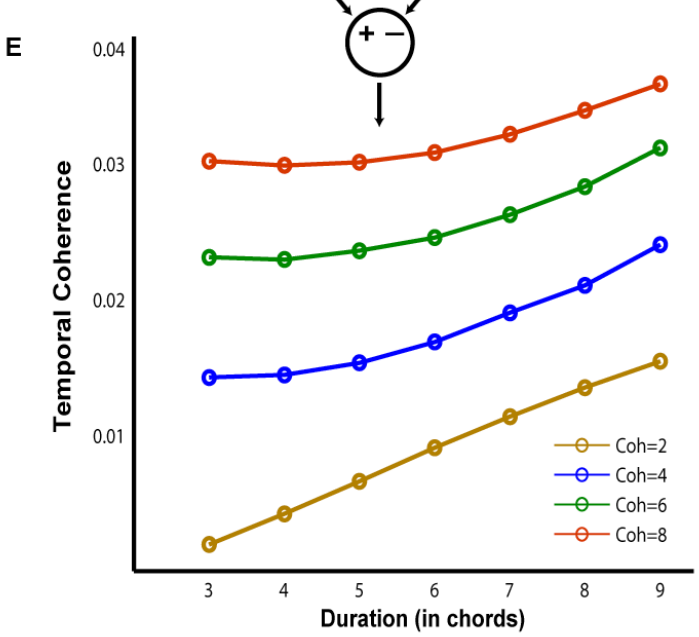
Isolated: expt. 5

- **Output:** Average cross-correlation$_{(figure)}$ - Average cross-correlation$_{(ground)}$



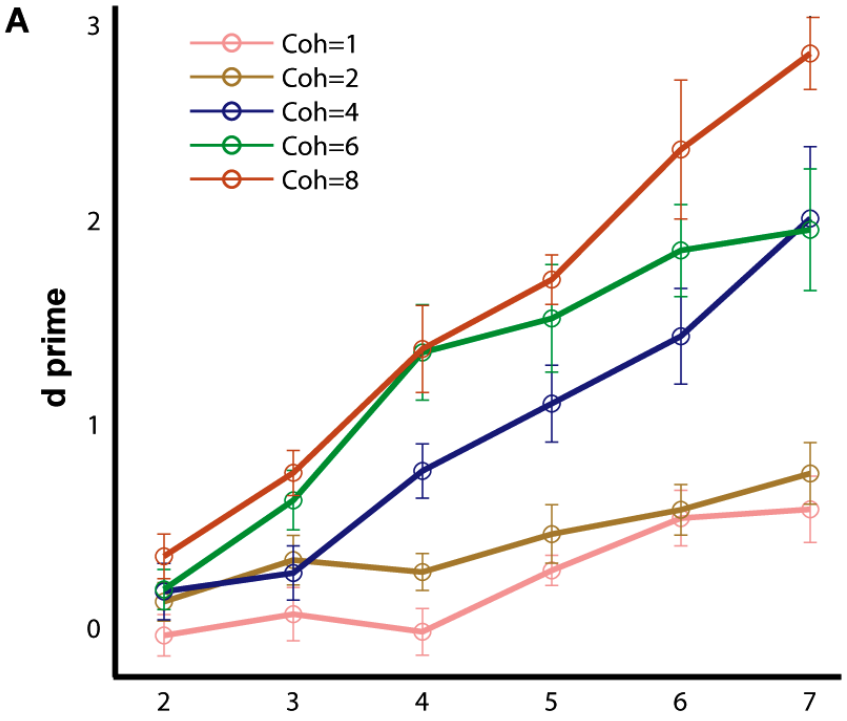
$\text{Coherence}_{\text{Figure}} = 1/N_{\text{iter}} * \sum\text{Max(coherence)}$

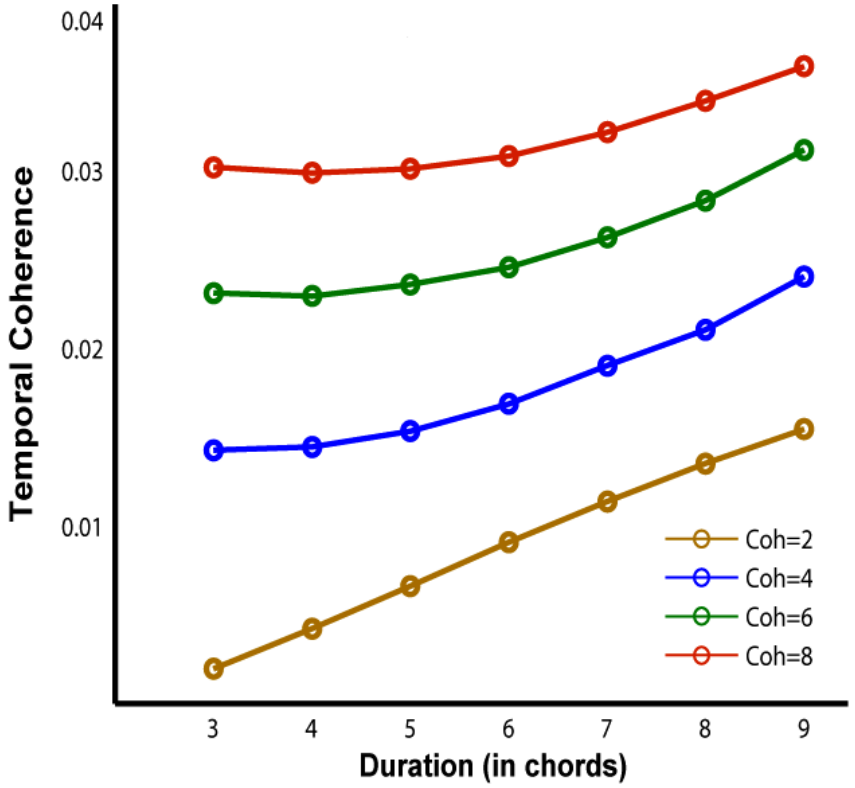
$\text{Coherence}_{\text{Ground}} = 1/N_{\text{iter}} * \sum\text{Max(coherence)}$



Isolated: expt. 5

IV. Behaviour vs. Temporal coherence





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.

Overall summary

SFG stimulus:

Represents a complex acoustic scene and allows parametric stimulus control
Listeners can segregate figure from ongoing background very well

Psychophysics:

Adaptation does not prove to be critical for segregation in SFG stimuli

Temporal coherence model:

Can explain figure-ground segregation in complex acoustic scenes as well

Acknowledgments

Deborah, Sukhbinder and Tim

Newcastle Auditory Group

Maria, Aiysha and Nicolas

UCL Ear Institute

Shihab Shamma

University of Maryland, College Park

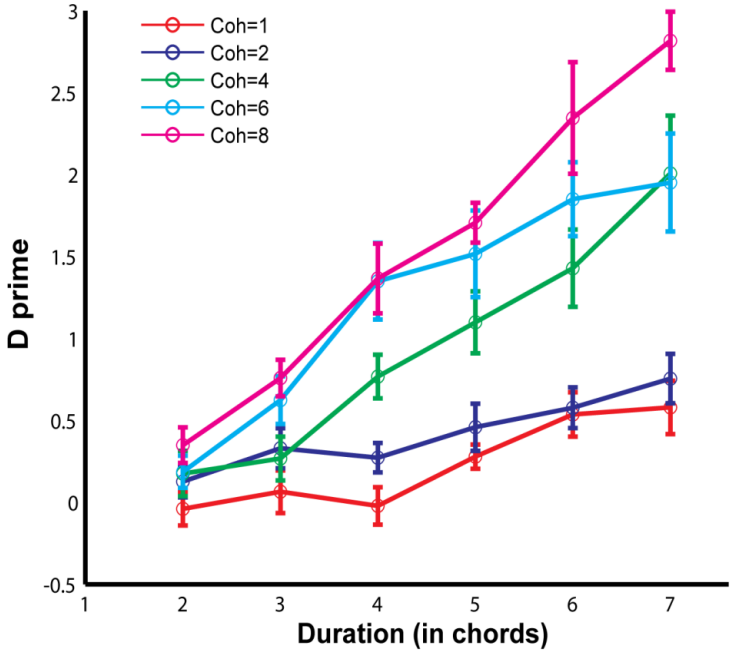
wellcome trust

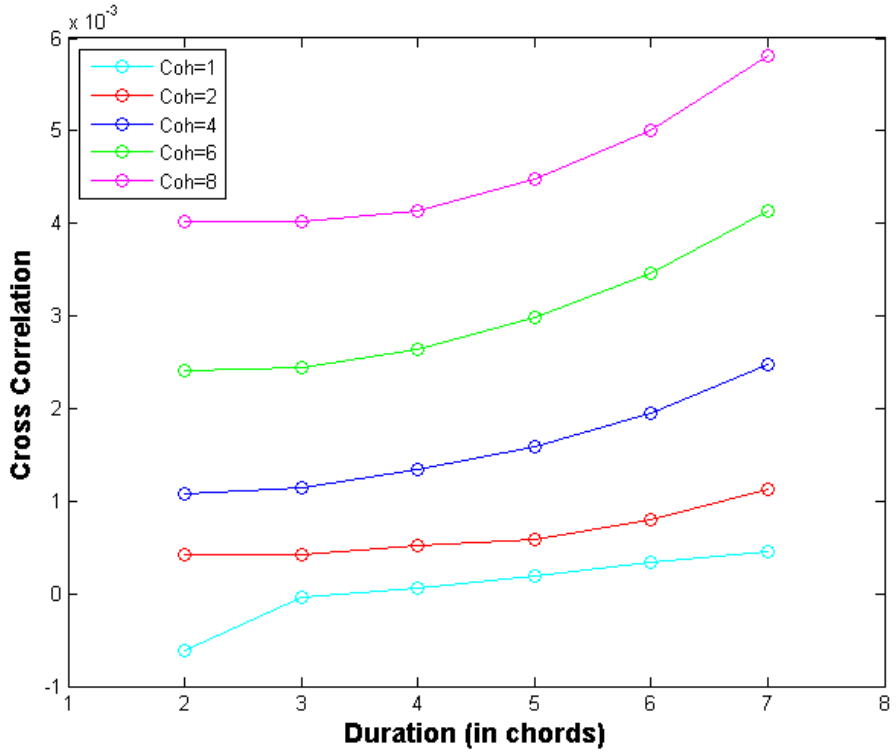

WOMEN
KINGFISHER
KINGFISHER

Questions ?

Expt. 1: Baseline (50 ms)

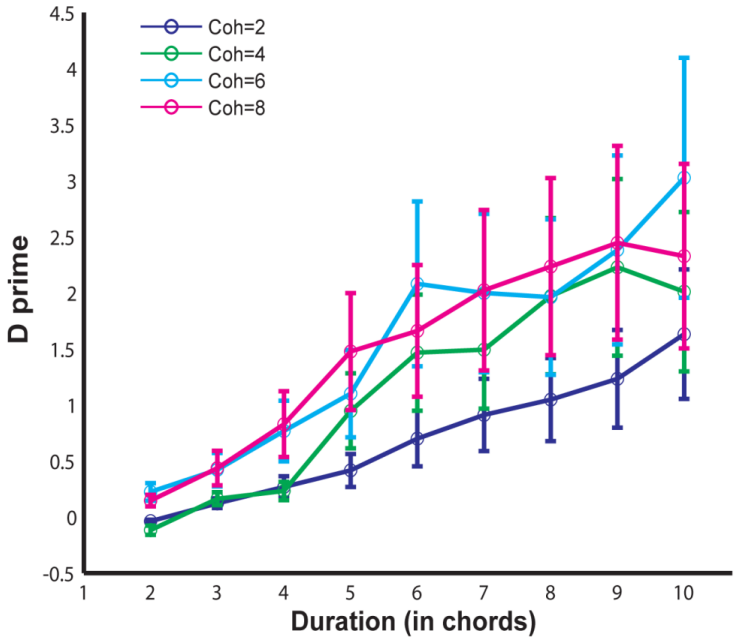
sv=8 cyc/oct, rv=20 Hz

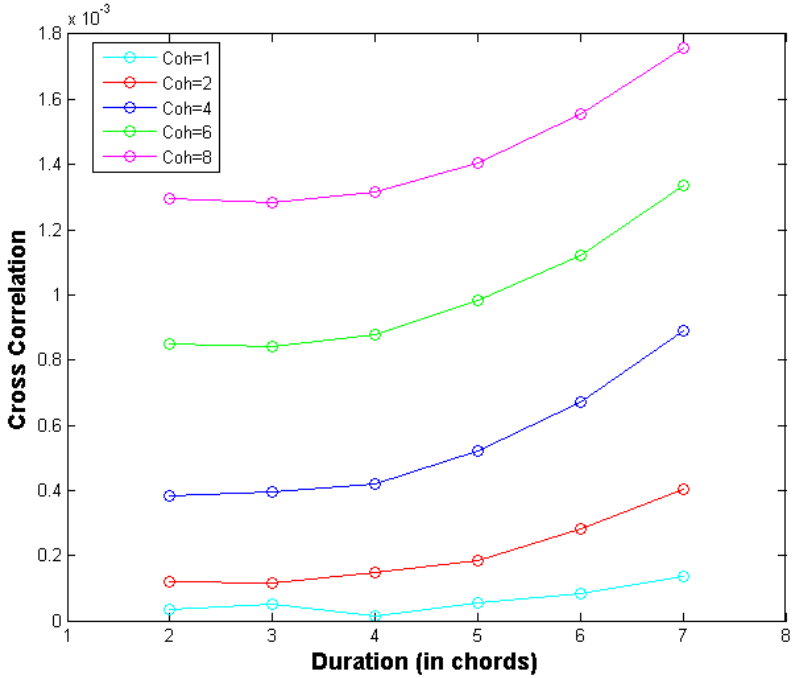




Expt. 2: Baseline (25 ms)

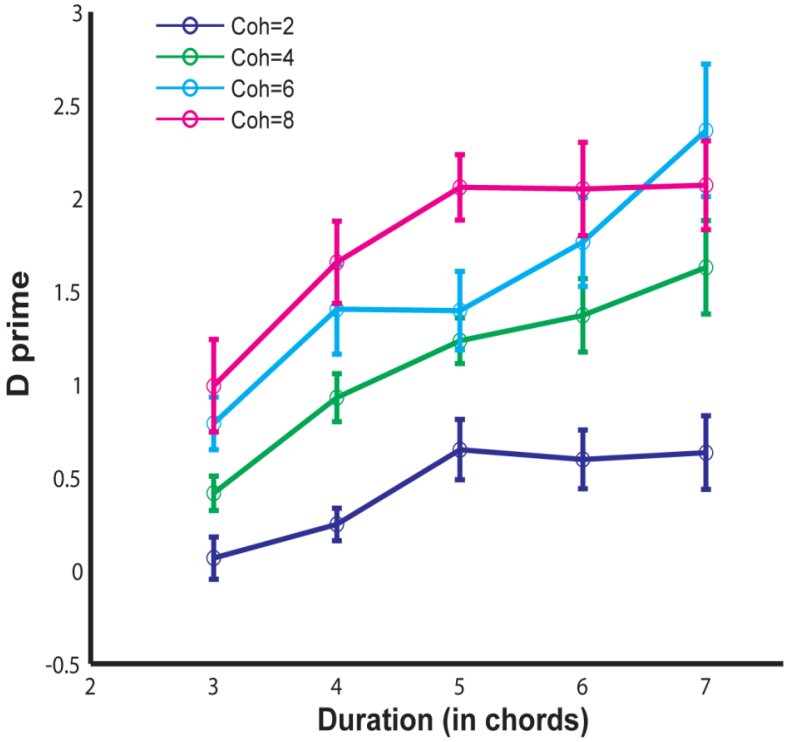
sv=8c/o, rv=40 Hz

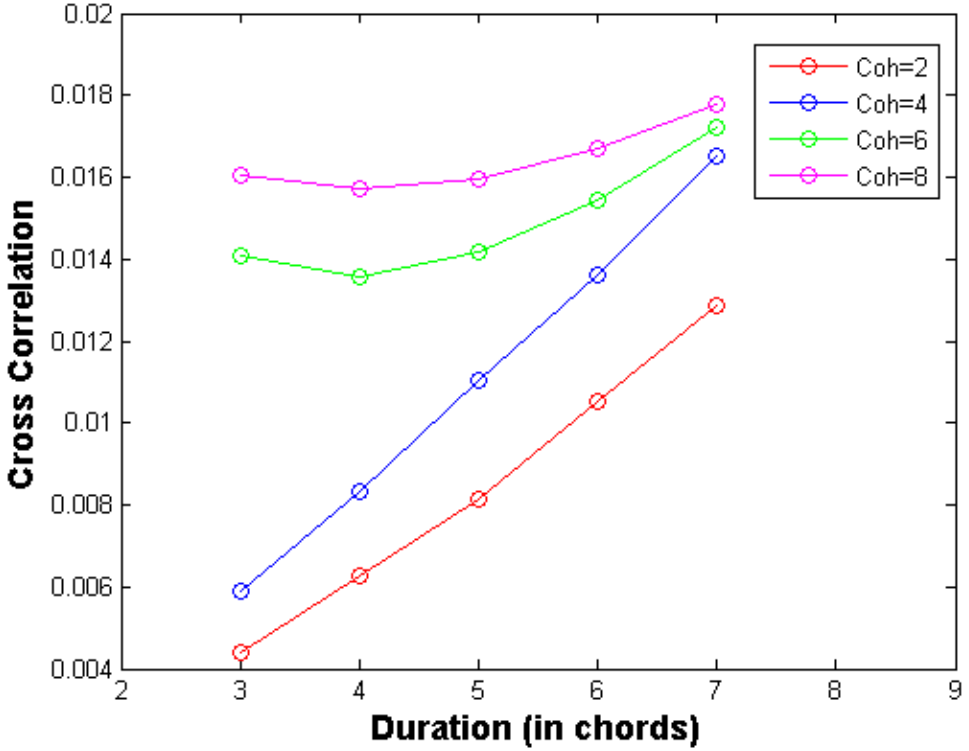




Expt. 3: SFG/Noise

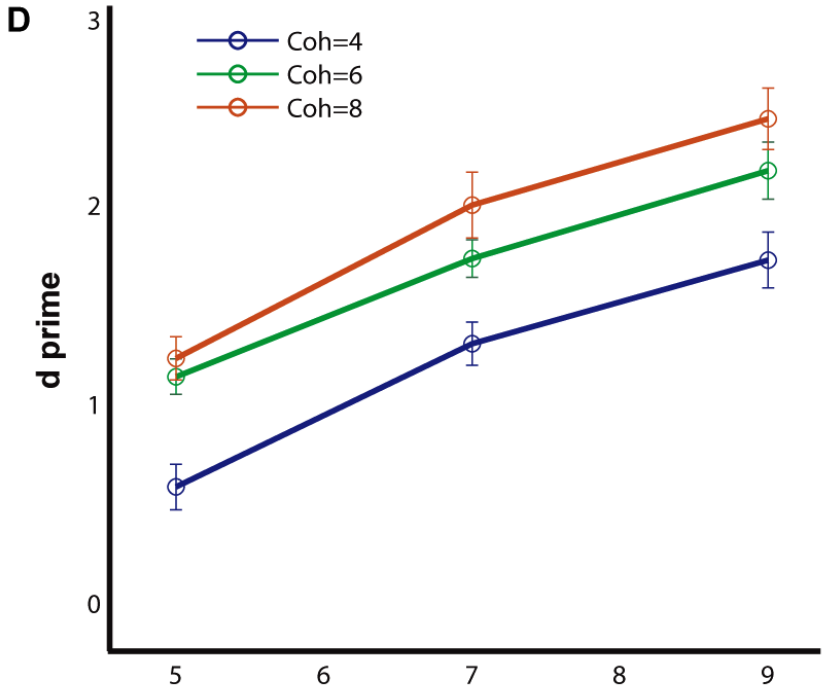
sv=24 Hz, rv=20 Hz

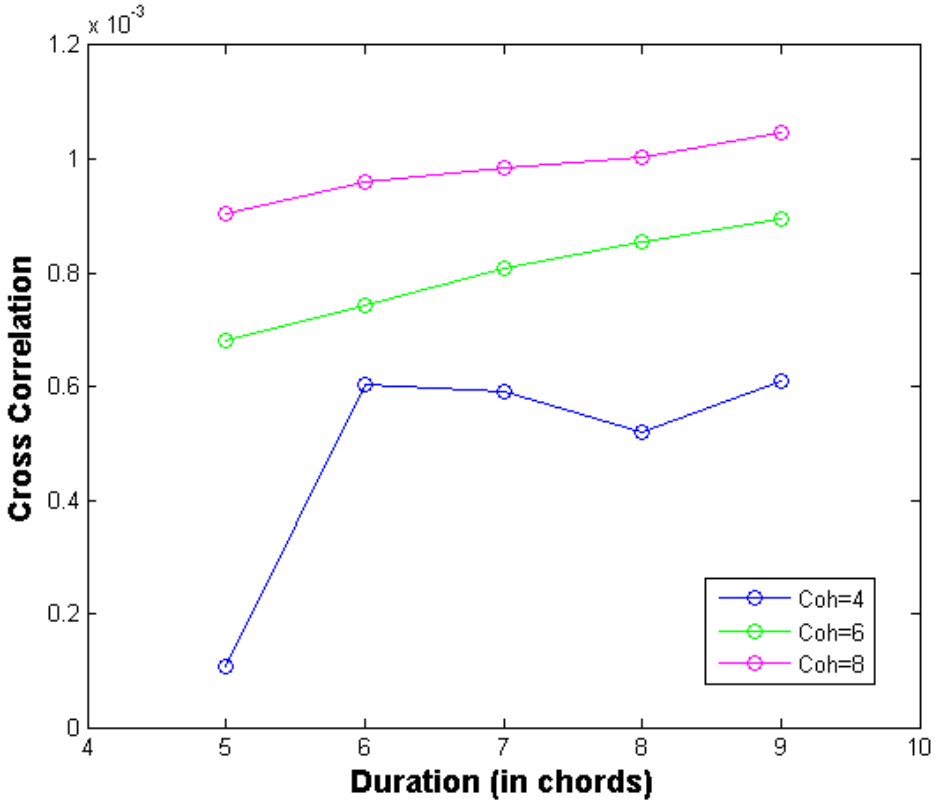




Expt. 4a: Ramps 2

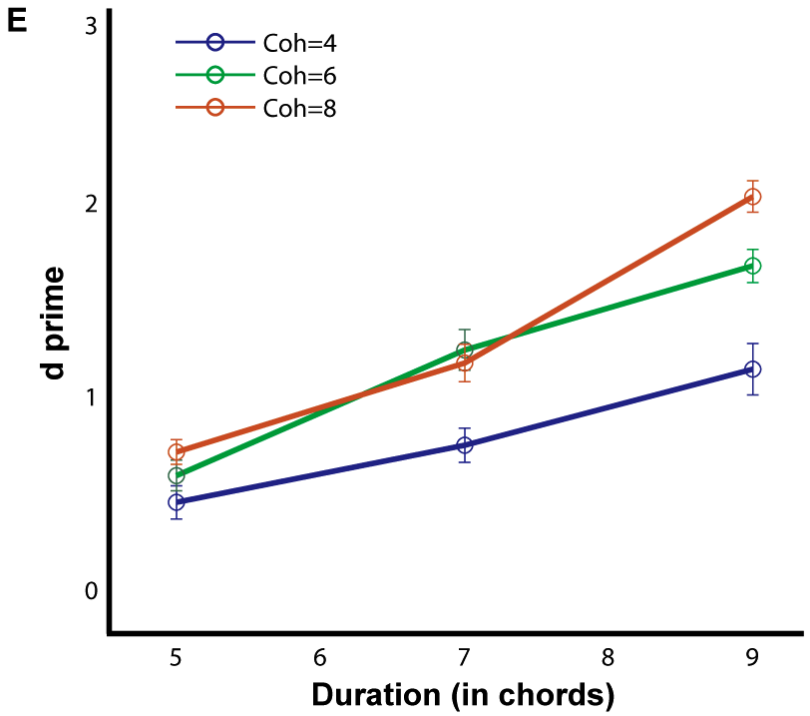
sv=8 c/o, rv=20 Hz

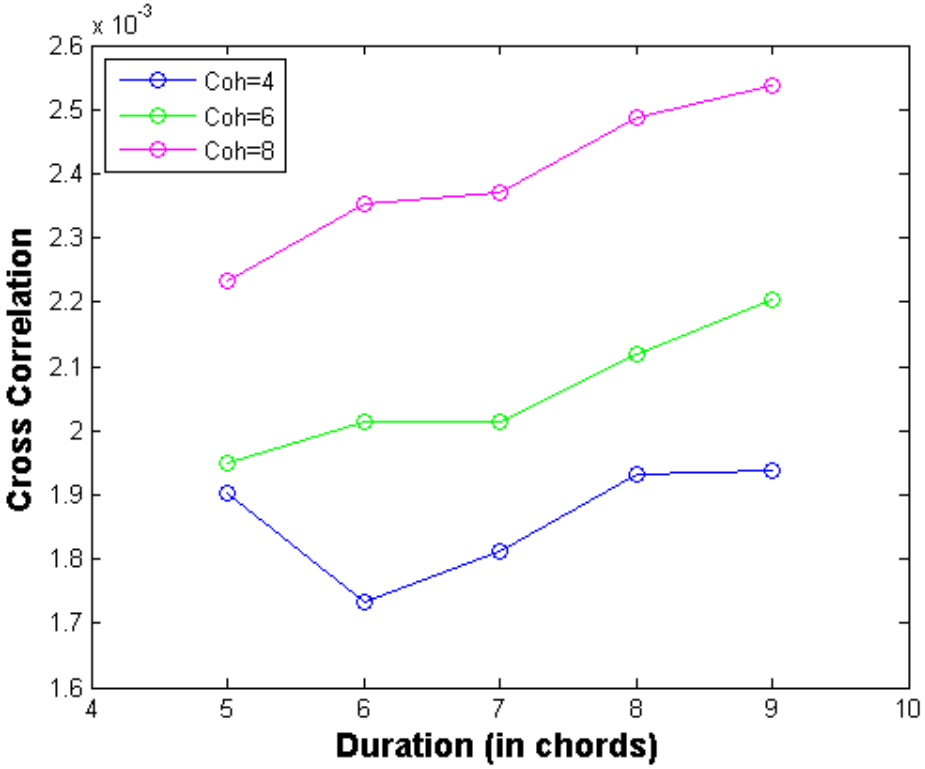




Expt. 4b: Ramps 5

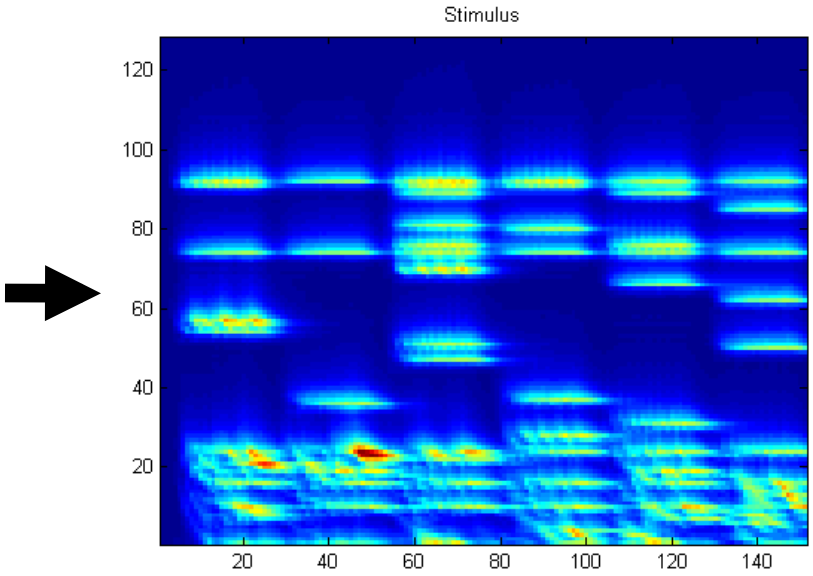
sv=4 c/o, rv=20 Hz

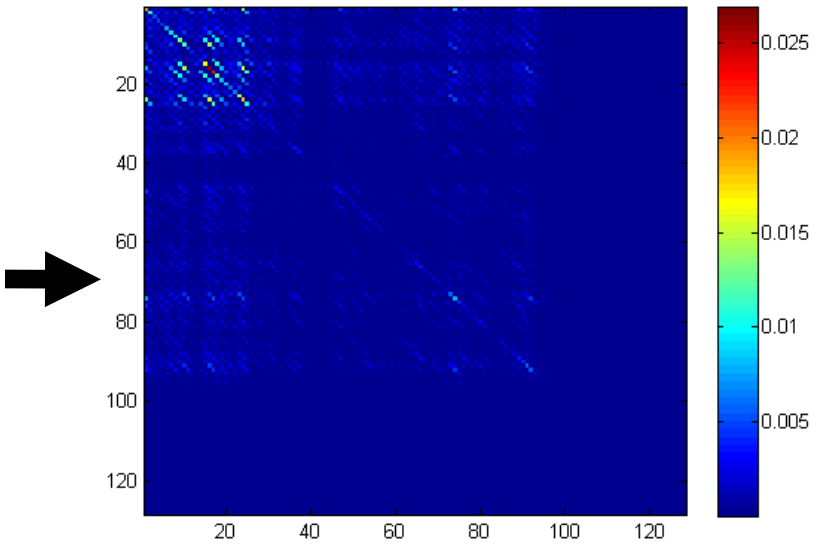




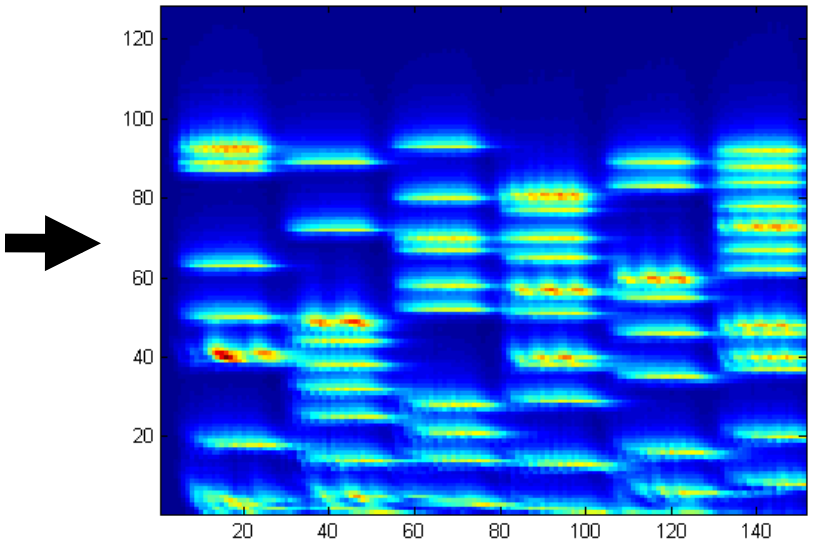
Isolated: expt. 5

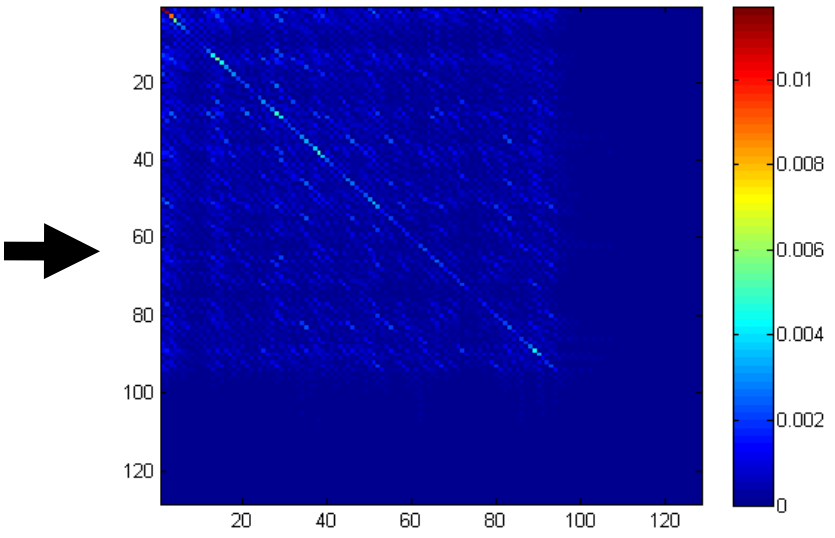
FIGURE





GROUND

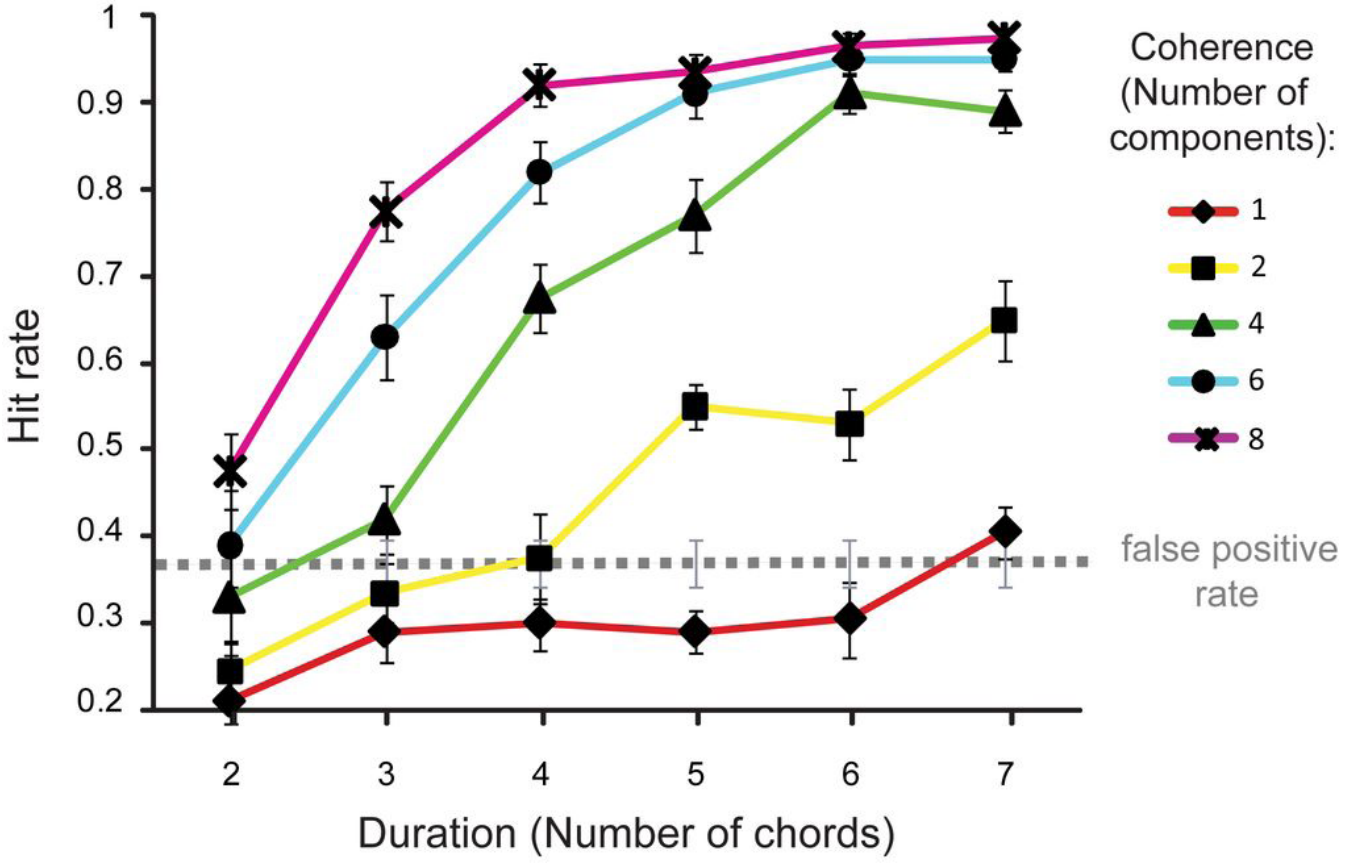




fMRI study

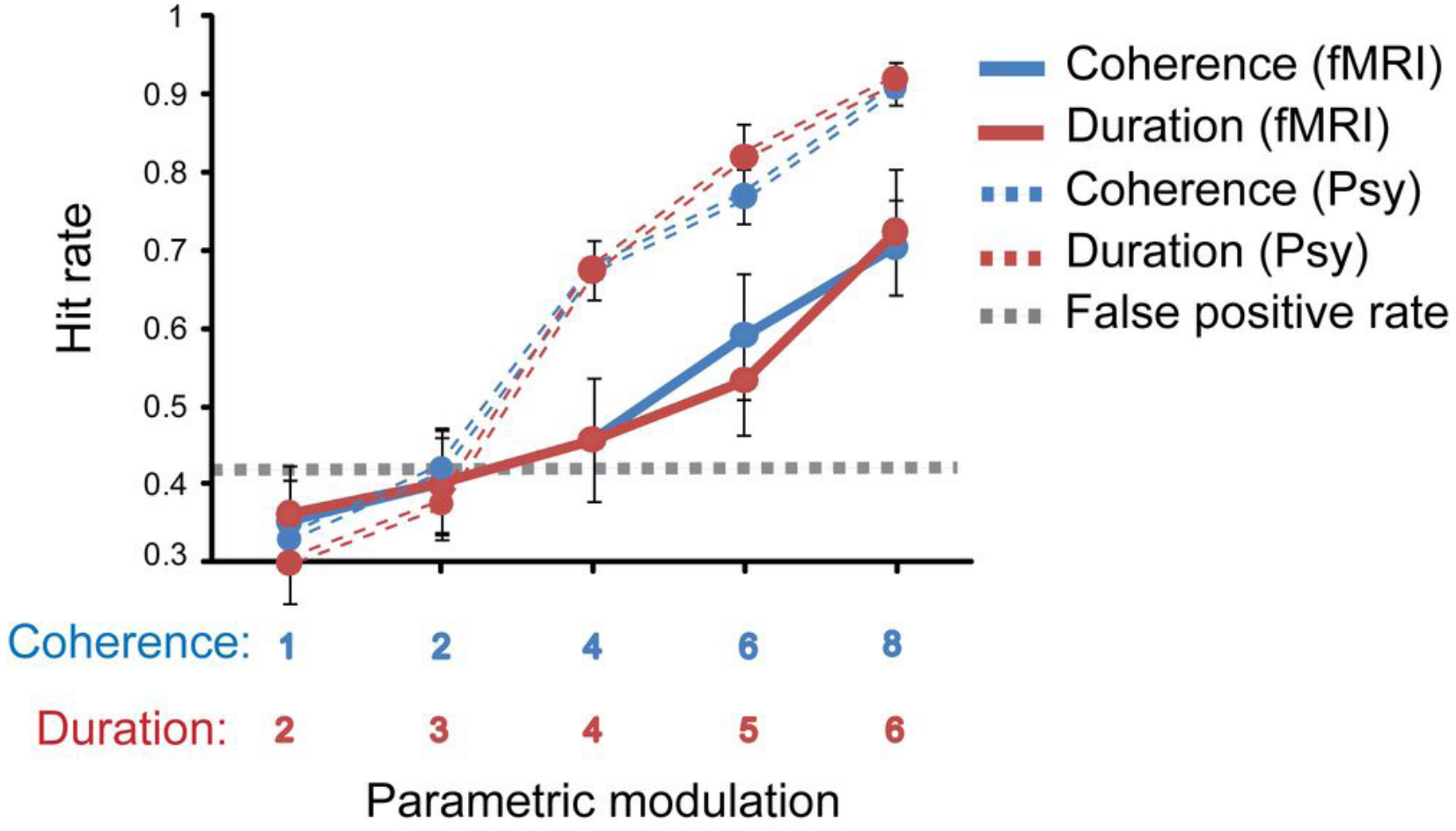
Behaviour in quiet

n=10



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

Behaviour in scanner



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

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

Figure (fixed coh: 4, dur: 2:7)

Figure (fixed dur: 4, coh: 1,2,4,6,8)

Background

Decoy

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

fMRI analysis

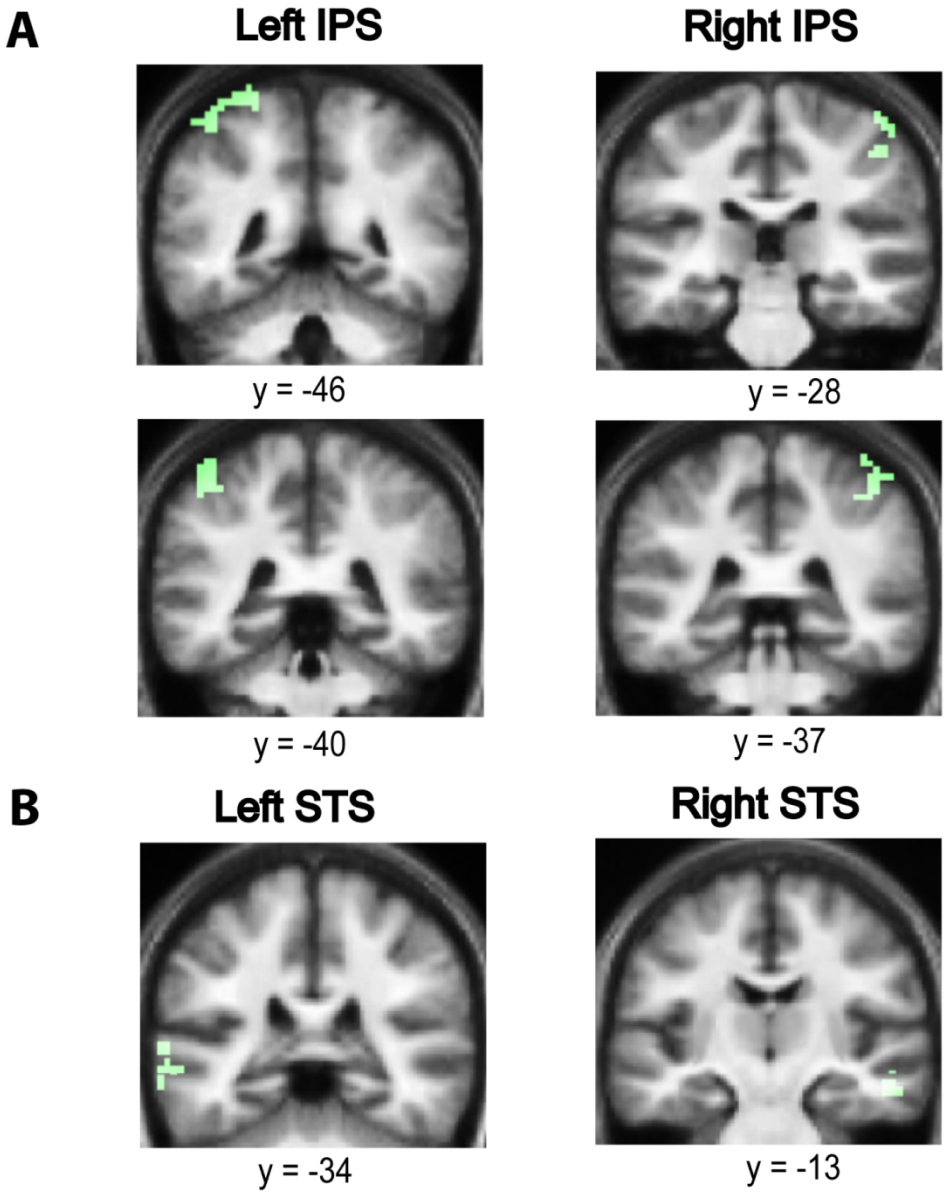
- 14 subjects (normal hearing, no audiological disorders)
- Standard pre-processing with SPM8
- Whole brain analysis
- Statistical model based on General Linear Model
- Random effects design

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)

Effects of Duration

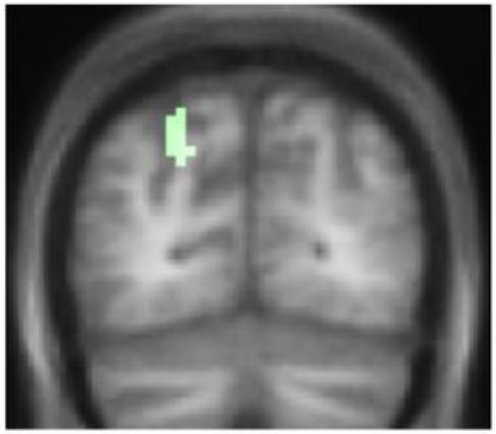


Also:
Bilateral MGB
Right PT

Effects of Coherence

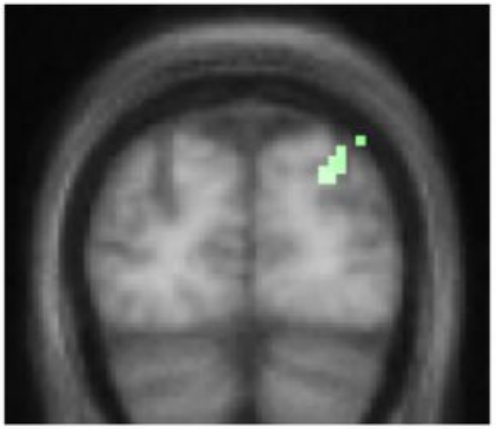
A

Left IPS



y = -73

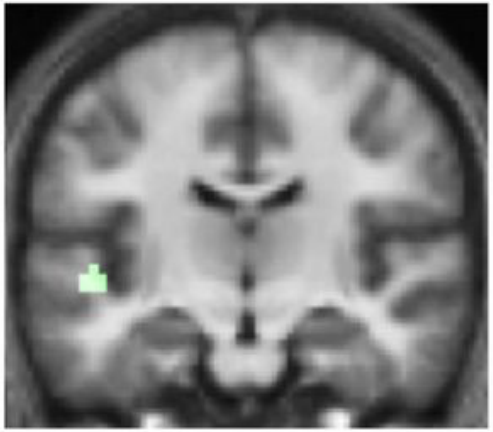
Right IPS



y = -82

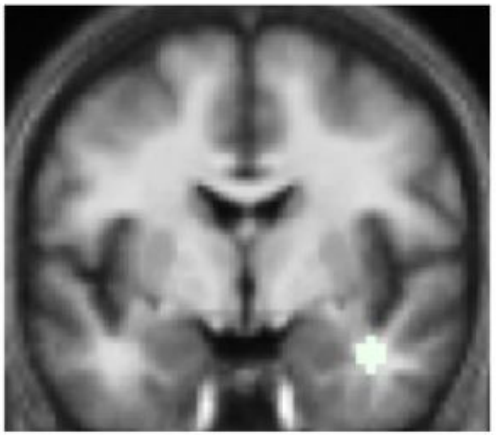
B

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?

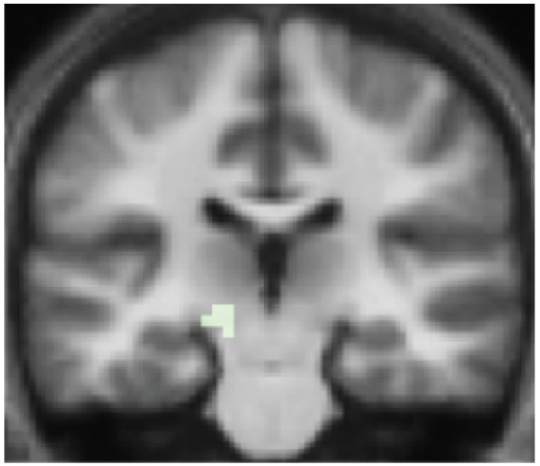
Teki, Chait et al., J Neurosci (2011)

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

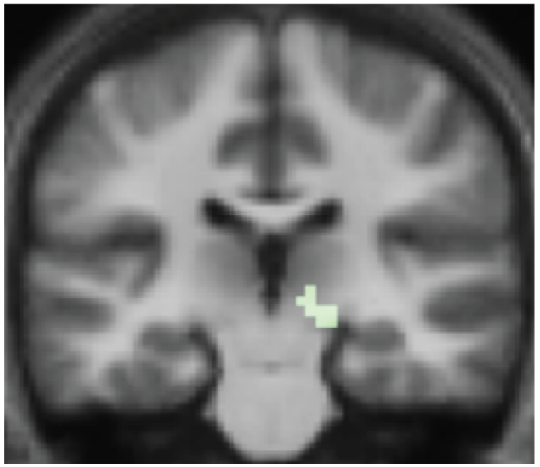
Local maxima for effects of duration, coherence as well as combined effects of duration and coherence are shown. Results are thresholded at $p<0.001$ (uncorrected)

Left MGB



y = -40

Right MGB



y = -37

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: $F(1,18) = 21, p < 0.001$