Competing streams at the cocktail party:
A neural and behavioral study of auditory attention

Computational Sensory Motor Laboratory

Center for Auditory and Acoustic Research

Juanjuan Xiang[1], Mounya Elhilali[2], Shihab A. Shamma[3,4], Jonathan Z. Simon[3,5]

[1] Starkey Laboratories Inc., Eden Prairie, MN. [2] Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore MD.
[3] Department of Electrical and Computer Engineering, University of Maryland, College Park, MD. [4] Institute for Systems Research, University of Maryland, College Park, MD. [5] Department of Biology, University of Maryland, College Park, MD.

University of Maryland 1856

Abstract

The mechanism by which a complex auditory scene is parsed into separate objects depends critically on attentional processes. We illuminate these mechanisms in a simultaneous behavioral-neurophysiological study, in which we manipulate subjects' attention to one of two competing rhythmic streams of an auditory scene. Our experimental results reveal that attention to the target stream correlates with a sustained increase in the neural target representation, as measured by magnetoencephalography (MEG), beyond auditory attention's well-known transient effects on onset responses. This enhancement, in both power and phase, occurs exclusively at the frequency of the target rhythm, for either stream, and is only revealed when contrasting two attentional states that direct subjects' focus to different features of the acoustic stimulus. The enhancement originates in auditory cortex and covaries with behavioral state, with a right-biased hemispheric asymmetry. Furthermore, for the slower stream, whose rhythmic rate is commensurate with that of speech prosody, the target's perceptual detectability improves over time, correlating strongly, within subjects, with the target representation's neural buildup.

Motivation & Methods

Attention

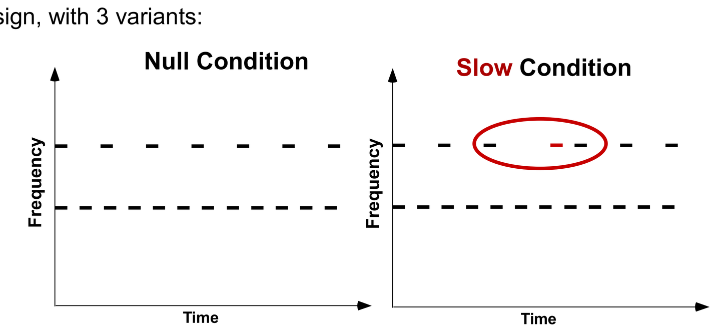
Cognitive process underlying our ability to focus on specific components of the environment while ignoring all others.

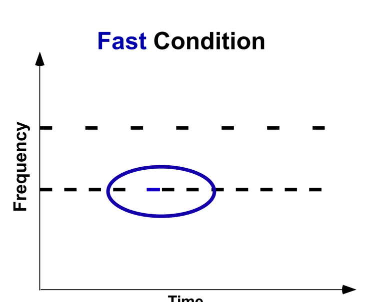
Setting

- In a cocktail party setting, it is critical to be able to enhance the percept of one auditory stream, while at the same time suppressing others.
- What is the contribution of attention to auditory scene analysis and what is its neural manifestation?

Paradigm

- Stimulus design, with 3 variants:





* Target rates: 4Hz, 7 Hz
* Stimulus duration: 5.25, 6.25, or 7.25 s
* ISI: 2.8–3.2 s
* Tone duration: 75 ms (incl. 10 ms cosine ramps)
* Spectral Rove Range: 250–500 Hz
* Spectral Separation: ±8 semitones
* Intensity Adjustment: approx. ISO-226 equal loudness
* Deviant shift (uniformly over duration): ±70 ms (4 Hz), ±40 ms (7 Hz)

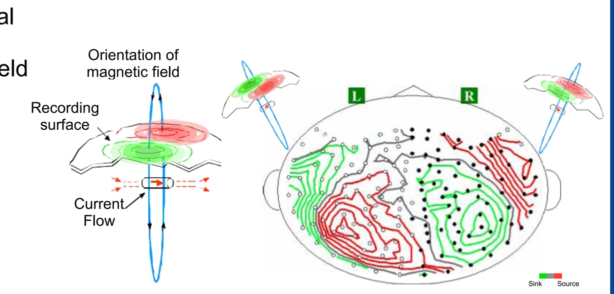
- Subjects perform two tasks in separate blocks of *identical* stimuli:
  - *Slow task*: detect temporal jitter in the slow (4 Hz) stream (present in 1/3 of stimuli)
  - *Fast task*: detect temporal jitter in the fast (7 Hz) stream (present in 1/3 of stimuli)

➡ **Contrast effects of attentional modulation to identical stimuli under two different tasks**

Technique

Magnetoencephalography + Psychophysics:
- Simultaneously acquired neural and behavioral data
- 26 subjects, performing both tasks
- Each task: 2 blocks of (3 conditions x 12 exemplars)
- Task order counterbalanced across subjects

Advantages of MEG:
- Non-invasive procedure, excellent temporal resolution of about 1 ms
- Not hemodynamic - measures magnetic field generated by neuronal current flow



References

Bidet-Caulet A, Fischer C, Besle J, Aguera PE, Giard MH, Bertrand O (2007). *Effects of Selective Attention on the Electrophysiological Representation of Concurrent Sounds in the Human Auditory Cortex*. Journal of Neuroscience, 27(35), 9252.

Bregman AS (1990) *Auditory scene analysis: The perceptual organization of sound*. MIT Press.

Elhilali M, Xiang J, Shamma SA, Simon JZ (2007) *Foreground and background at the cocktail party - Interaction between attention and auditory pop-out*, Society for Neuroscience abstracts.

Hillyard SA, Hink RF, Schwent VL, and Picton, TW (1973). *Electrical Signs of Selective Attention in Human Brain*. Science 182, 171-180.

Micheyl C, Shamma SA, Oxenham AJ (2007) "*Hearing out repeating elements in randomly varying multi-tone sequences: A case of streaming?*", Chapter in "Hearing - from basic research to applications", Springer.

Naatanen R, Tervaniemi M, Sussman E, Paavilainen P and Winkler I (2001). *"Primitive intelligence" in the auditory cortex*. Trends in Neurosciences 24, 283-288.

Simon JZ and Wang Y (2005) *Fully Complex Magnetoencephalography*, J. Neurosci. Methods. 149(1), 64-73.

Srinivasan R, Russell DP, Edelman GM, Tononi G (1999). *Increased Synchronization of Neuromagnetic Responses during Conscious Perception*. Journal of Neuroscience, 19(13), 5435.

Acknowledgments

Support provided by NIH grants 1R01DC007657, 1R01DC008342, and (via the CRCNS NSF/NIH joint mechanism) 1R01AG027573. We thank David Poeppel for comments and discussion. We are grateful to Jeff Walker for excellent technical support.

1. Neural Responses

Power:

Top: Power spectral density of the 4 Hz neural response component for a single subject in Slow (left) and Fast (right) tasks, averaged over twenty channels.

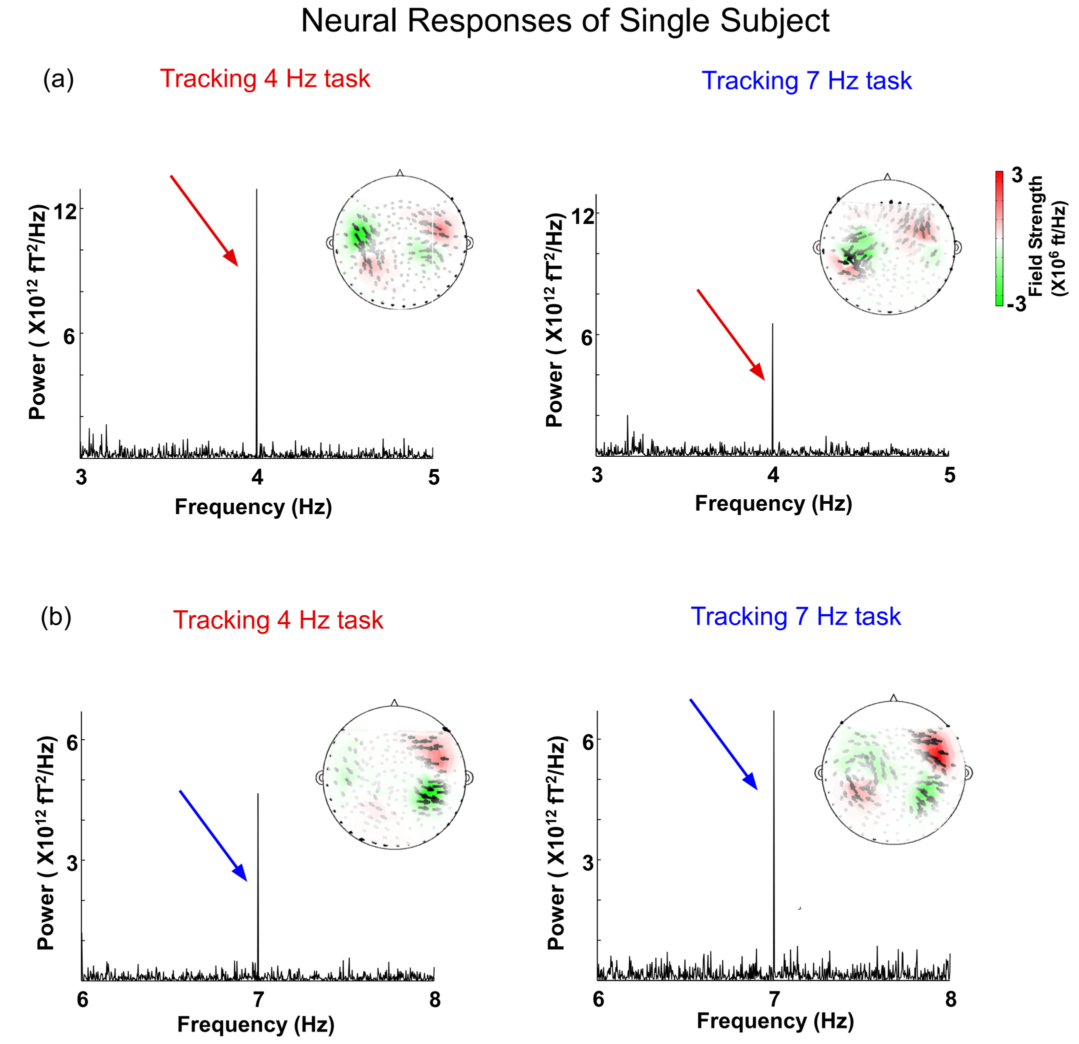
Bottom: Power spectral density of the 7 Hz neural response component for the same subject in Slow (left) and Fast (right) tasks.

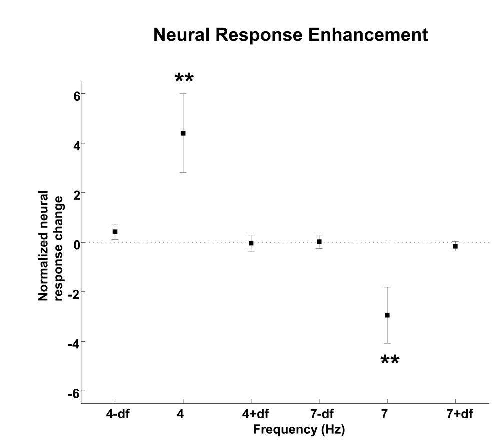
Insets: The MEG magnetic field distributions of the peak response component. Red and green contours represent the target magnetic field strength projected onto a line with constant phase.

Slow task: strong neural signal component at 4 Hz
vs.
Fast task: weaker neural signal component at 4 Hz

Slow task: weaker neural signal component at 7 Hz
vs.
Fast task: strong neural signal component at 7 Hz

Neural activity (of target rhythm) originates in auditory cortex—specifically 19 ± 6 mm anterior to the location of planum temporale/M100-source, consistent with Heschel's Gyrus, the site of core auditory cortex.





Normalized neural response is the ratio of the neural response power at target rate to the average power of the background neural activity, averaged over the 20 best channels. Error bars represent standard error.

Normalized neural responses difference between tasks shows enhancement exclusively at target rates.

Phase Coherence:

Left: Phase coherence difference between tasks for 4 Hz neural response component.

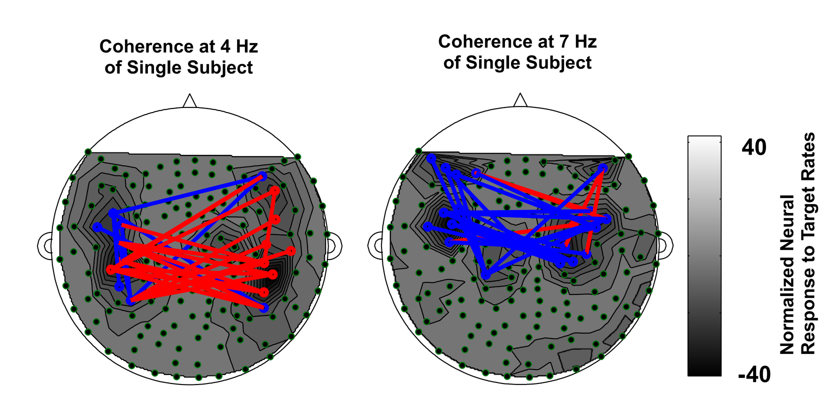
Right: Phase coherence difference between tasks for 7 Hz neural response component.

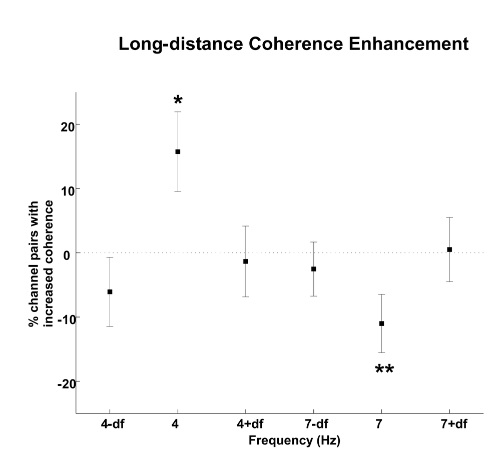
Channel pairs with robust coherence change at target rates for single subject.

Increased coherence = Blue
Decreased coherence = Red

Connections are overlaid on the contour map of normalized neural response at target rates.

Channel pairs with robust coherence change are distributed both within and across hemispheres.





Phase coherence is computed as the ratio of average cross spectrum between each channel pair, and average power spectrum of the individual channels (Srinivasan et al., 1999). The difference between number of channel pairs with robust increased coherence and channel pairs with decreased coherence is normalized over the total number of long range channel pairs. Error bars represent standard error.

Phase coherence difference between between tasks shows enhancement exclusively at target rates.

Note:

The physical stimulus in both cases is the same
=> task-specific attentional influence.

This attentional effect on the neural signal is not just momentary but is sustained over the duration of stimulus.

Behavioral Performance:

Slow task: d-prime: 2.9
Fast task: d-prime: 1.8

The extent of modulatory effects of top-down attention may be related with the natural saliency of stimuli (4 Hz stream more salient than 7 Hz).

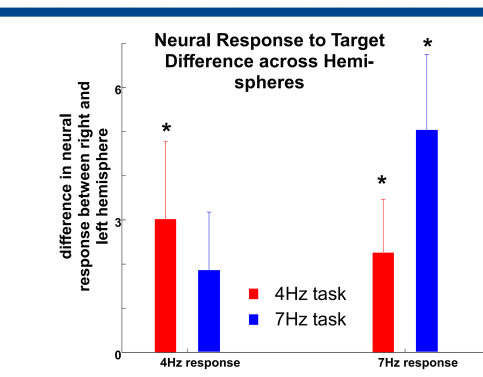
➡ Auditory attention strongly modulates the sustained neural representation of the target. This neural representation is located at the level of sensory auditory cortex.

2. Hemispheric Asymmetries

Task-dependent Hemispheric Asymmetry:

Right hemisphere shows a greater normalized neural response than left at (attended) target rates.

For unattended target rates, the right hemisphere dominance is observed at 7 Hz, but not at 4 Hz.



20 channels with strongest normalized neural response at target rates are chosen from each hemisphere. Neural responses averaged across channels are subtracted across hemispheres for each task and averaged over subjects. Error bars represent standard error.

3. Responses build up over time

Tracking the Slow Stream:
Perceptual detectability of target increases over time

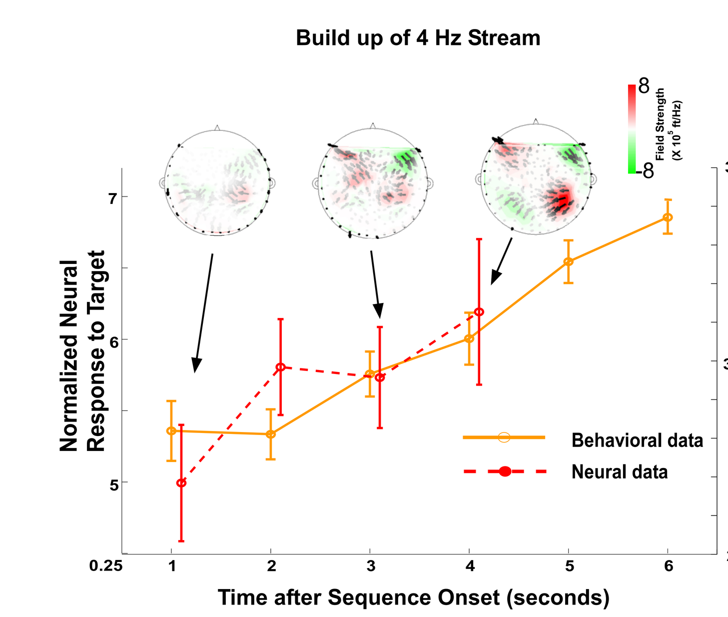
Trend suggests that mechanisms for target detection is mediated by processes conjectured to play a role in object formation (consistent with previous findings of build-up of auditory streaming).

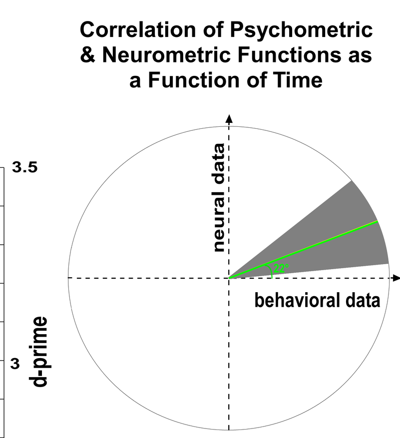
Time-course of this behavioral buildup is strongly correlated with an increased neural representation of the target over time.

Buildup of neural responses over time is seen only when integrated over several periods of the target rhythm.

Correspondence between the neural and behavioral temporal buildups is confirmed using bootstrap.

Behavioral buildup, neural buildup, and strong correlation are all consistent with that seen in an earlier study (Elhilali et al., 2007) in which a 4 Hz auditory stream is embedded in a noisy background.



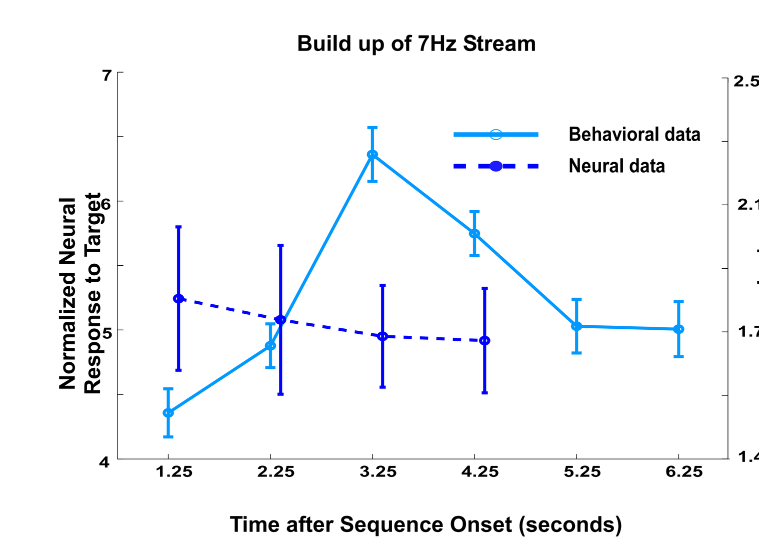


Neural responses at target rate were divided temporally: analysis epochs were divided into 4 segments of 1 s duration (starting at 1.25 s post stimulus) and corresponding segments were concatenated across epochs. The segment duration 1000 ms is commensurate with both 4 Hz and 7 Hz.

Tracking the Fast Stream:
Perceptual detectability of target increases, but then decreases, over time. Even the peak detectability is much less than that in the Slow Stream tracking case.

Time-course of the neural representation of the target is largely flat, showing neither build-up, nor correlation with the perceptual detectibility.

Buildup of neural responses over time is not seen even when integrated over several periods of the target rhythm.



➡ **Even though the sensory target signal is unchanged, attention allows its neural representation to grow over time, following dynamics strongly correlated with the time-course of its perceptual buildup. But why only for the slower stream?**

Discussion

Auditory attention strongly modulates the sustained neural representation of the target (complementing well-known transient attentional effects). This neural representation is located at the level of sensory auditory cortex.

This study allows us to monitor the evolution in time of attentional processes as they interact with the sensory input, and demonstrates that the neural representation of a target signal that also follows the same temporal profile of the buildup based on listeners' detectability performance. This buildup effect suggests the implication of coherent or synchronous neural activity as a neural mechanism of selective attention.

These findings support a view of a tightly coupled interaction between the lower level neural representation and the higher level cognitive representation of auditory objects, in a clear demonstration of the cocktail party effect.