Foreground and background at the cocktail party: A neural and behavioral study of top-down and bottom-up auditory attention

Computational Sensory Motor Laboratory

Center for Auditory and Acoustic Research

Mounya Elhilali[1], Juanjuan Xiang[2], Shihab Shamma[1,2], Jonathan Simon[2,3]

1 Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD 21218, USA.
2 Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20815, USA.
3 Department of Biology, University of Maryland, College Park, MD 20815, USA.

University of Maryland 1856

Abstract

The mechanism by which a complex auditory scene is parsed into coherent objects depends on poorly-understood interactions between task-driven and stimulus-driven attentional processes. We illuminate these interactions in a simultaneous behavioral-neurophysiological study, in which we manipulate subjects' attention to different features of an auditory scene (with a regular target embedded in an irregular background). Our experimental results reveal that target focused attention correlates with a *sustained* increase in the neural target representation, beyond auditory attention's well-known transient effects. This enhancement, in both power and coherence, occurs not between states of attending and non-attending, but rather between two separate attentional states, each focusing on different acoustic features of the stimulus. The enhancement originates in auditory cortex and covaries with both the behavioral state appropriate to the task and the bottom-up saliency of the target. Furthermore, the target's perceptual detectability improves *over time*, correlating strongly with the target representation's *neural buildup*. These results have substantial implications for models of foreground/background organization and mechanisms mediating auditory object formation.

Motivation & Methods

Attention

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

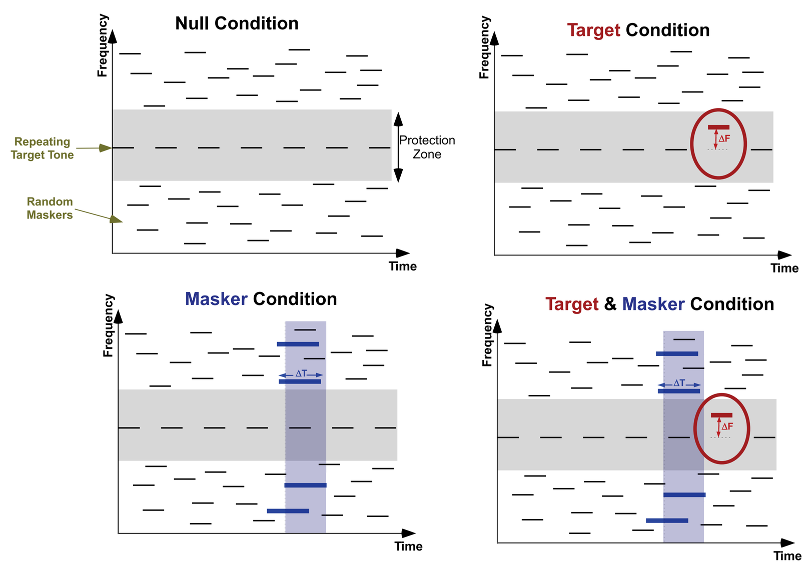
- Attention can be **bottom-up** (sound-based) or **top-down** (task-dependent). Both processes are thought to operate in conjunction in order to selectively process sensory information, and pass the relevant cues to higher auditory and cognitive areas.

Setting

- In a cocktail party setting, the extraction of a foreground from the background (e.g. signal from noise) can be thought of as a multifaceted process that draws on bottom-up gestalt primitives, as well top-down control including attention and memory.
- What is the contribution of attention to auditory scene analysis and what is its neural manifestation?

Paradigm

- Stimulus design commonly used in **Informational Masking** experiments, with 4 variants:



- Target rate: 4Hz
- Stimulus duration: 5.5 s
- Tone duration: 75 ms
- Target shift: +/- 2 semitones
- Masker elongation: 400 ms
- Target roved [250-500] Hz

- Subjects perform two tasks in separate blocks:
  * *Target task*: detect frequency shift (ΔF) in repeating target signal;
  * *Masker task*: detect sudden temporal elongation (ΔT) of masker notes

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

Technique

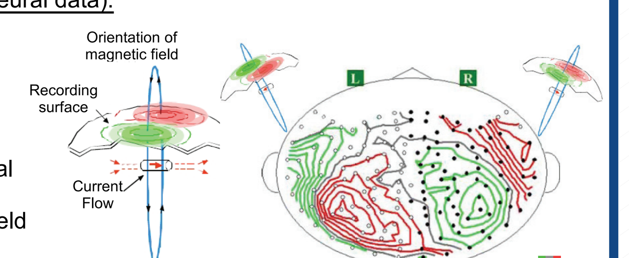
Psychoacoustics:

- 9 subjects, performing both tasks
- soundproof room, sounds dichotically presented over headphones
- subjects interacted with a Graphical User Interface
- Each task: 180 stimuli (3 protection zones x 4 conditions x 15 exemplars)
- Subjects self-paced between trials, no feedback was provided

Magnetoencephalography (behavioral and neural data):

- 14 subjects, performing both tasks
- Each task: 3 blocks of (1 protection zone x 4 conditions x 15 exemplars)

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



References

Bregman AS (1990) Auditory scene analysis: The perceptual organization of sound. MIT Press.
Fishman YI, Reser DH, Arezzo JC, Steinschneider M (2001) Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. Hear. Res. 151, 167-187.
Hillyard, S.A., Hink, R.F., Schwent, V.L., and Picton, T.W. (1973). Electrical Signs of Selective Attention in Human Brain. Science 182, 171-180.
Maunsell JHR, Treue S (2006) Feature-based attention in visual cortex. Trends in Neurosciences, 29, 317-322
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, J. Z. and Y. Wang (2005) Fully Complex Magnetoencephalography, J. Neurosci. Methods. 149(1), 64-73.

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

1. Behavioral Performance

Target Task:

Detectability of regular tone becomes easier with increased protection zone

Influence of protection zone is consistent with the notion that the frequency selectivity of neurons in the central auditory system is an important determinant of stream segregation

Masker Task:

Same manipulations of protection zone do not substantively affect masker task performance

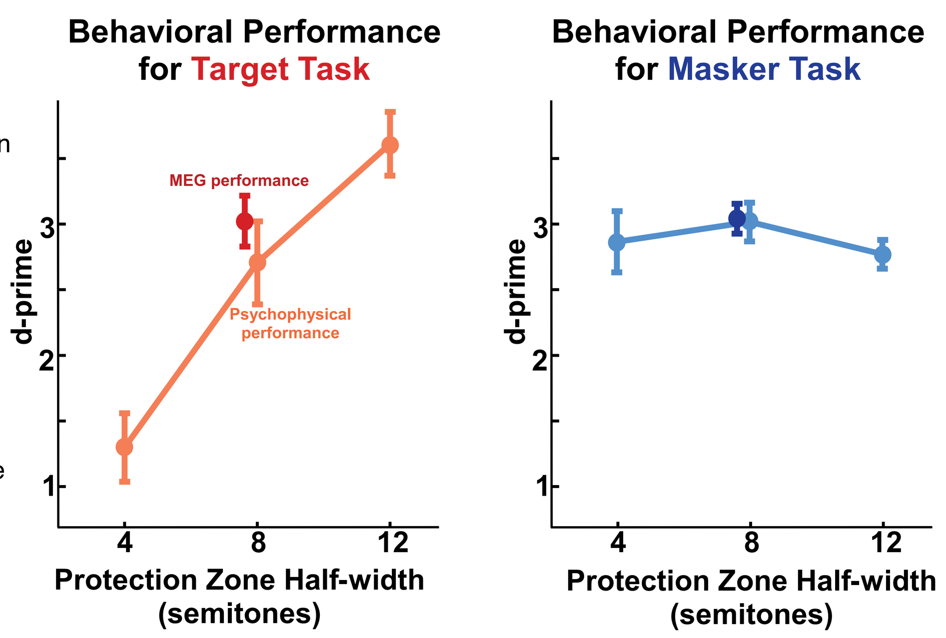
The masker task, designed to divert attentional resources away from the target, involves a more diffuse attention

Compared to target task, it reflects different top-down bias in the way the same stimulus is parsed.

Note:

Behavioral performance during MEG and psychophysical testing (8 st) are the same

At 8 st, performance is comparable between target and masker tasks (d-prime ~ 3).
=> comparable attentional load?



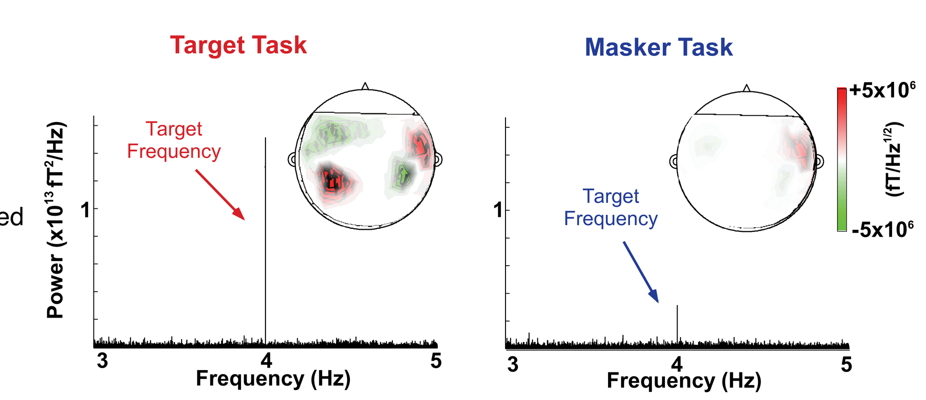
➡ **The experimental design contrasts selective attention in two tasks of comparable difficulty, involving attending to different components of the same, identical stimulus.**

2. Neural Responses

Target task: strong 4Hz component in neural signal,

Masker task: response entrained at 4Hz noticeably suppressed

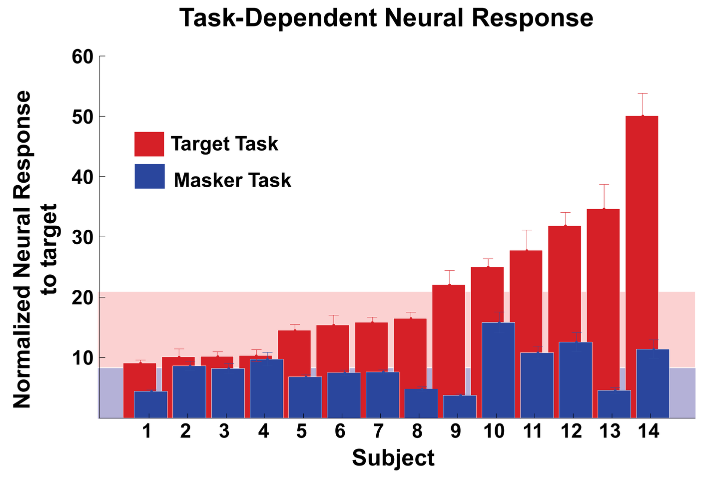
Neural activity (of target rhythm) originates in auditory cortex



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



Additional Results

Attentional power enhancement effect is consistent across subjects.

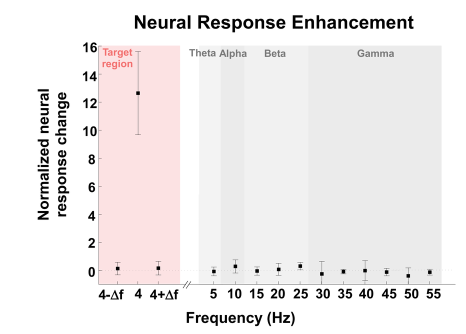
(11 out of 14 subjects: statistically significant difference between tasks)

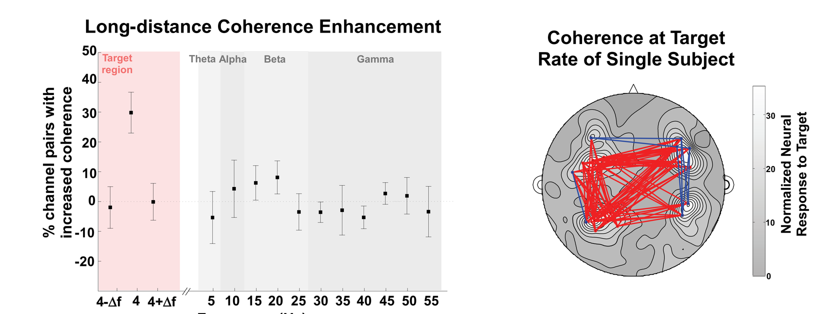
Attentional power enhancement effect occurs only at target rhythm rate

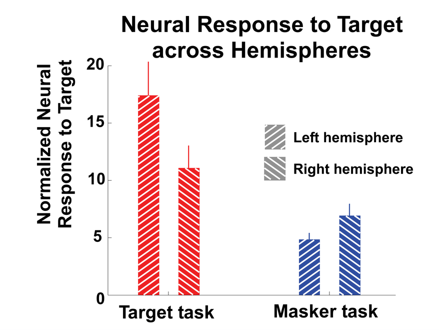
Attentional phase coherence enhancement effect occurs only at target rhythm rate

No direct correlation between target task neural response and d-prime

Task-Dependent hemispheric asymmetry in power enhancement effect







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

3. Effect of Bottom-up saliency

Frequency of the target note affects its audibility (bottom-up saliency effect)
=> audibility difference of ~ 5dB over [250-500] Hz

Target Task:

Behavioral data confirms target task is more salient
=> increased subject performance (d-prime) for high-frequency (>350Hz) relative to low-frequencies

Correlated with this trend is an increased neural power of target frequency for high vs. lower frequencies

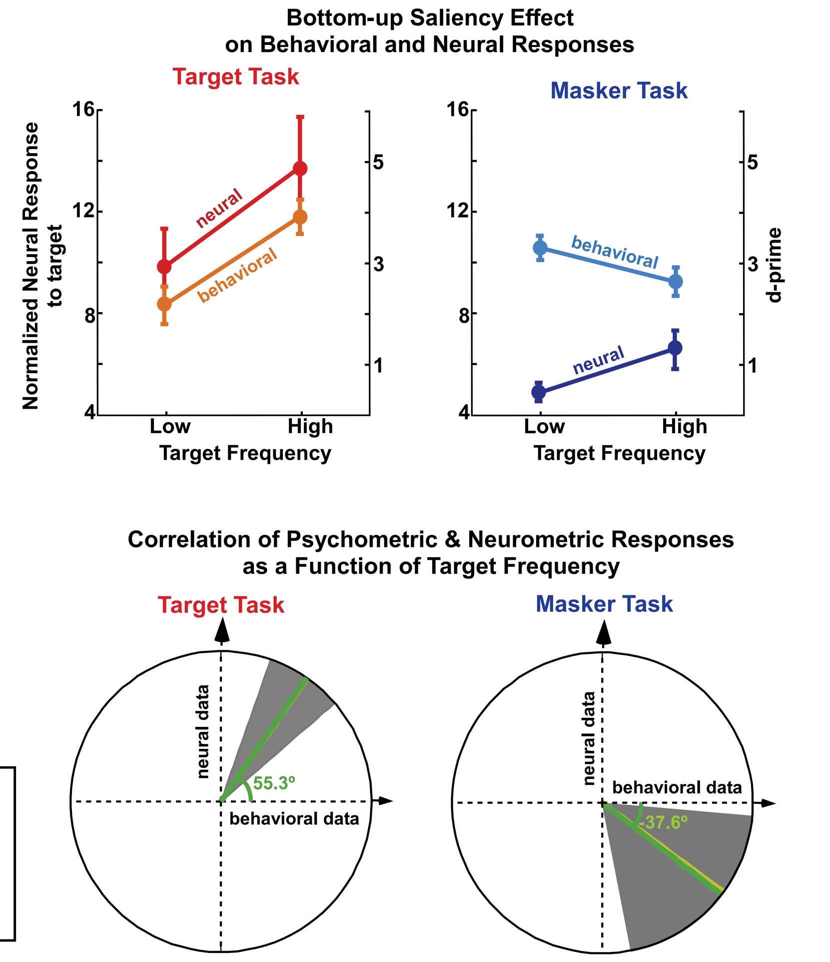
Masker Task:

Neural power is increased for high-frequency target reflecting their increased audibility

For more prominent targets, subjects' performance of the background task deteriorates indicating a distraction/interference effect

Psychometric vs. Neurometric data:

Neural vs. behavioral correspondence is confirmed using bootstrap
[angle/slope between neural signal and d-prime per-subject]



➡ **Together with the behavioral demands of the task, the bottom-up saliency of the target note shapes both neural and behavioral responses.**

4. Responses build up over time

Target Task:

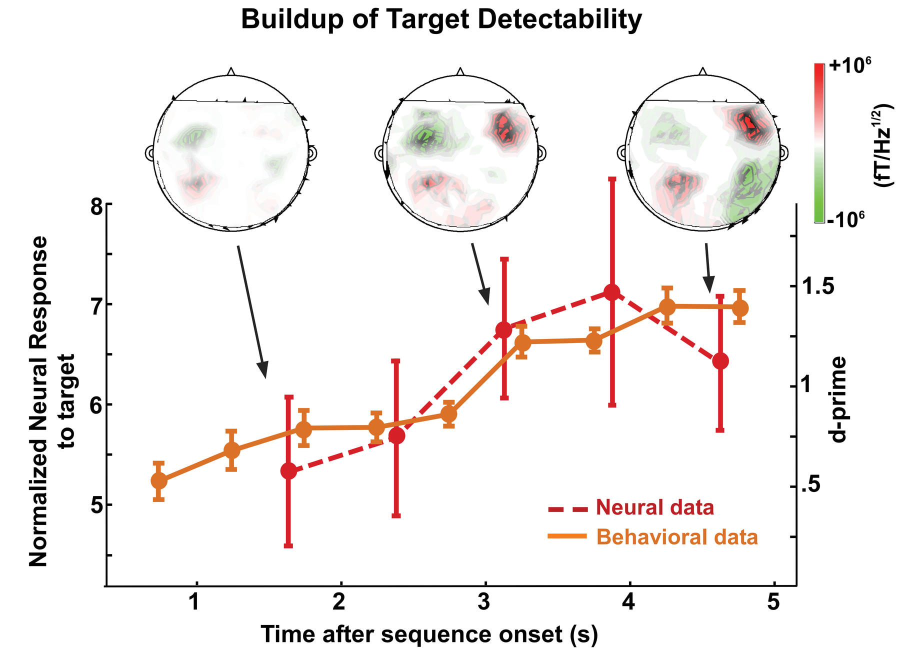
Perceptual detectability of target increases over time

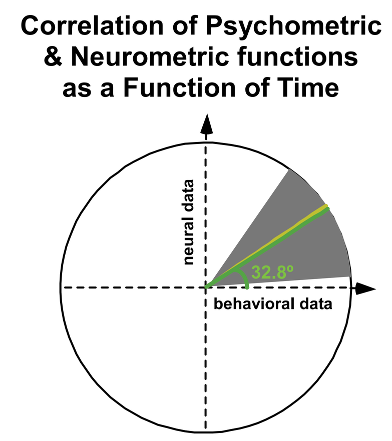
Trend suggests that mechanisms for target detection is mediated by processes conjectured to play a role in object formation (consistent with previous findings of build-up of auditory streaming)

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

Psychometric vs. Neurometric data:
Correspondence between the neural and behavioral temporal buildups is confirmed using bootstrap





➡ **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.**

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.

The enhanced acoustic saliency, which causes an increase in perceptual detectability, also correlates with an increase in the sustained neural signal

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.