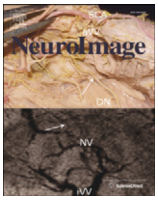
NeuroImage 84 (2014) 681–687

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg



Switching auditory attention using spatial and non-spatial features recruits different cortical networks

Eric Larson [a], Adrian K.C. Lee [a,b,*]

[a] Institute for Learning and Brain Sciences, University of Washington, 1715 Columbia Road N, Box 357988, Seattle, WA 98195, USA
[b] Department of Speech & Hearing Sciences, University of Washington, 1715 Columbia Road N, Box 357988, Seattle, WA 98195, USA

ARTICLE INFO

Article history:
Accepted 26 September 2013
Available online 3 October 2013

Keywords:
Auditory attention
Magnetoencephalography
Electroencephalography
Temporoparietal junction
Inferior parietal supramarginal part
Spatial attention
Pitch processing

ABSTRACT

Switching attention between different stimuli of interest based on particular task demands is important in many everyday settings. In audition in particular, switching attention between different speakers of interest that are talking concurrently is often necessary for effective communication. Recently, it has been shown by multiple studies that auditory selective attention suppresses the representation of unwanted streams in auditory cortical areas in favor of the target stream of interest. However, the neural processing that guides this selective attention process is not well understood. Here we investigated the cortical mechanisms involved in switching attention based on two different types of auditory features. By combining magneto- and electro-encephalography (M-EEG) with an anatomical MRI constraint, we examined the cortical dynamics involved in switching auditory attention based on either spatial or pitch features. We designed a paradigm where listeners were cued in the beginning of each trial to switch or maintain attention halfway through the presentation of concurrent target and masker streams. By allowing listeners time to switch during a gap in the continuous target and masker stimuli, we were able to isolate the mechanisms involved in endogenous, top–down attention switching. Our results show a double dissociation between the involvement of right temporoparietal junction (RTPJ) and the left inferior parietal supramarginal part (LIPSP) in tasks requiring listeners to switch attention based on space and pitch features, respectively, suggesting that switching attention based on these features involves at least partially separate processes or behavioral strategies.

Introduction

The ability to flexibly switch attention between competing auditory stimuli based on task demands is critical to communication in many settings. Directing attention improves detection of relevant signals (Posner et al., 1980), biases sensory cortices (Petkov et al., 2004; Wu et al., 2007; Yantis et al., 2002) even prior to relevant stimulus onset (Voisin et al., 2006), and enhances encoding of preferred stimuli (e.g., Mesgarani and Chang, 2012). However, the top–down control signals that allow listeners to “tune into” a stream of interest are not fully understood.

Recent neuroimaging studies suggest that top–down auditory attentional control engages multiple neural mechanisms prior to target stimulus processing (Hill and Miller, 2010; Lee et al., 2013). Areas including inferior frontal gyrus and the superior temporal sulcus (STS) show greater activity in preparing to attend to stimuli based on pitch compared to left premotor areas for attending based on spatial features. Additionally, tasks that involve speech processing or working memory engage parietal areas including the left inferior parietal supramarginal part (LIPSP; Hutchinson et al., 2009; Zheng et al., 2013). These types of studies suggest that directing auditory attention engages a distributed cortical network with the involvement of different areas changing depending on the features of interest.

However, it remains less clear how attention can be flexibly switched between multiple, simultaneous auditory streams at will. Endogenous attentional control has been studied using cognitive tasks (Kiesel et al., 2010), but most auditory attention switching studies have focused on stimulus-driven attention (Driver and Spence, 1998; Shinn-Cunningham, 2008). Neuroimaging evidence suggests that a dorsal cortical network mediates switching attention hypothesized to operate supramodally (Corbetta et al., 2008). This idea is supported by behavioral similarities of attention-switching costs in audition and vision (Koch et al., 2011), and fMRI studies showing multiple cortical regions active in both visual and auditory tasks (Shomstein and Yantis, 2006; Wu et al., 2007). One recent auditory M-EEG study also shows that the right temporoparietal junction (RTPJ) and premotor areas are more active prior to sound onset when switching attention between spatially separated sounds (Larson and Lee, 2013). In this study, a visual cue prompted subjects to switch spatial attention immediately prior to

---

Abbreviations: RTPJ, Right temporoparietal junction; LIPSP, left inferior parietal supramarginal part; M-EEG, Magneto- and electroencephalography; STS, superior temporal sulcus.

* Corresponding author at: University of Washington, 1715 NE Columbia Road, Box 357988, Seattle, WA 98195, USA. Fax: +1 206 221 6472.

E-mail addresses: larsoner@uw.edu (E. Larson), akclee@uw.edu (A.K.C. Lee).

1053-8119/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.neuroimage.2013.09.061

the onset of sound stimuli, making it challenging to separate the influences of exogenous visual cueing from those of switching auditory attention.

A compelling unresolved question is the extent to which these mechanisms of switching *spatial* attention generalize to *non-spatial* features. While evidence suggests that redirecting auditory attention in space shares the supramodal attention switching system employed by the visual system, this could be because audio-visual stimuli in natural settings tend to have co-varying locations—the concordance of spatial information across these two modalities thus makes for a natural sharing of attentional control mechanisms. Pitch cues in audition, however, do not have an immediate visual correlate in natural stimuli; moreover, the processing of pitch is known to involve distinct neural circuitry. In this study, we therefore sought to test the hypothesis that switching auditory attention based on spatial and non-spatial cues engages distinct underlying neural mechanisms. Here we investigate this hypothesis by using streams that differ in pitch but have no spatial differences, and by using streams with only different spatial percepts. To identify neural mechanisms involved in switching attention, we use anatomical MRI-constrained M-EEG measurements during a behavioral task that requires subjects to switch selective attention between two simultaneous auditory streams. Here we temporally separate the switch- or maintain-attention cueing from the period of time during which subjects can switch attention—this allows us to take advantage of the timing information in M-EEG to help separate the neural responses to cueing from those involved in goal-driven attention modulation.

Materials and methods

Subjects

Twelve healthy normal-hearing subjects participated in the experiment, each giving informed consent according to procedures approved by the University of Washington. All subjects had eyesight correctable to 20/20 with magnet-compatible glasses or contact lenses; had hearing within the normal audiological range in both ears (less than 20 dB HL from 250 Hz to 8 kHz at octave frequencies); were aged 19–31, with 2 female; had Edinburgh handedness scores 50–100 (except one left-handed subject with −95); excluding the left-handed subject had no discernible effect on results aside from decreasing statistical power in permutation tests by a factor of 2 (see Materials and methods). We also excluded two subjects due to data being too noisy for reliable analysis.

Behavioral task and stimuli

In each trial, subjects were presented with two simultaneous, competing auditory streams. Each of the two streams (the target to be attended, and masker to be ignored) consisted of six letters, with letters presented one every 400 ms such that the target and masker letter onsets were always concurrent (see Fig. 1). However, between the third and fourth letters, a 600 ms gap was inserted (effectively creating "triplets" of letters before and after the gap) to allow subjects time to switch attention between the two streams. The use of this gap period also facilitated appropriate comparison between the space-only and pitch-only conditions, since during this period the auditory input (no target or masker) was the same. For example, on a given trial a target stream could consist of the letters "EYR-JOQ," whereas the masker stream could consist of "ROJ-EME." Subjects were instructed to count the number of "E"s in only the target stream of letters and respond using a button box with the number of "E"s heard. The letters were presented simultaneously (with "E"s in both the target and masker stream) to help force listeners to selectively attend to only one of the two streams at a time in order to perform the task correctly (and subjective post-experiment polling suggests that this is the strategy subjects employed). The target and masker letter streams were monotonized, and were identical except for 1) the letters each stream contained (and the counts of "E"s in each stream), and 2) either an imposed inter-aural time delay (ITD) of ±300 μs or a pitch difference of 8.5 semitones (see Stimulus generation, below).

Subjects were instructed to attend to a visual fixation dot at the center of the screen throughout the experiment. At the start of each trial, subjects were first cued to attend to a stream by the presentation of two auditory "A" tokens processed in the same way as the target auditory stream (either a leftward or rightward ITD shift, or upward or downward pitch shift). These cue "A"s were then followed by a noise burst to disrupt any potential buildup of auditory streaming. Simultaneous with the onset of the first "A" cue, subjects were shown a visual diamond, X, or square that cued them to 1) maintain attention to the cued talker for all six target letters, 2) switch attention from the initial target to competing talker after the first three target letters, or 3) count the number of times the visual fixation dot flashed. This last condition was run to serve as a similar-task, auditory-unattended control condition for localizing auditory responses based on frequency tagging (see Stimulus generation). There was a 500 ms gap between the offset of the cue noise burst and the first target and masker letters, as cue-target intervals of over 500 ms typically allow for adequate task preparation (Meiran et al., 2000). In both the target and masker streams, there was a 600 ms gap period separating the first and last three letters to provide subjects sufficient time to switch attention. For example, on a given trial, assume a subject heard a high-pitched "AA" with a diamond cue (maintain-attention condition), and then were presented a high-pitched "EYR-JOQ" stream and a low-pitched "ROJ-EME" stream. The correct number of "E"s to report would be one, since there was one "E" in the high-pitched stream. However, if this same trial had been cued with an X (switch-attention condition), the correct number of

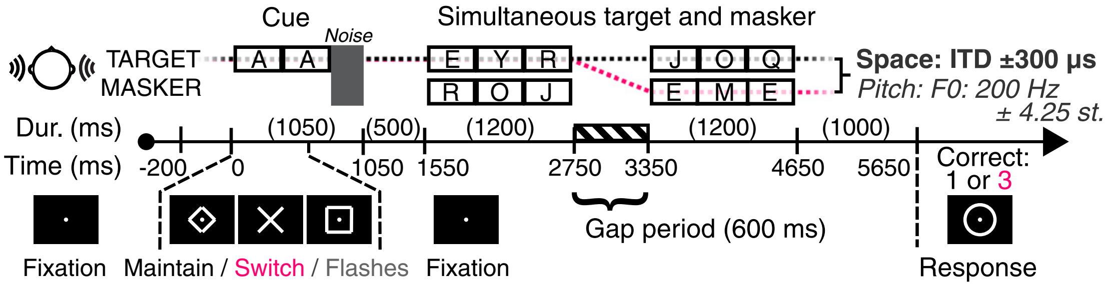


Fig. 1. Subjects were instructed to attend to one of two simultaneous auditory streams, maintaining or switching attention between them halfway through the streams. Concurrent auditory (two "A"s and a noise burst) and visual (600 ms duration diamond, X, or square) cues are followed by two simultaneous six-letter streams. Subjects report the number of target "E"s (0–3) once a circle response cue appears. In the example shown here, the correct response on a maintain-attention trial (black) would be one, while the correct response on a switch-attention trial would be three. Note that between the first three target letters (examples shown here) and the last three target letters, a 600-ms gap period was used, during which no target or masker auditory stimuli were played. During this gap (which is identical across conditions), subjects were to either maintain attention (diamond cue; black dotted line) or switch attention (X cue; pink dotted line) between the two streams. Target and masker auditory stimuli had either a spatial separation with equal pitches, or a pitch difference with equivalent spatial cues.