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

Eric Larson[1] and **Adrian KC Lee**[1,2]

Eric Larson: larsoner@uw.edu; Adrian KC Lee: akclee@uw.edu

[1]Institute for Learning and Brain Sciences, University of Washington, 1715 Columbia Road N, Box 357988, Seattle, WA 98195 USA

[2]Department of Speech & Hearing Sciences, University of Washington, 1715 Columbia Road N, Box 357988, Seattle, WA 98195 USA

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 electroencephalography (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.

Keywords

Auditory attention; magnetoencephalography; electroencephalography; temporoparietal junction; inferior parietal supramarginal part; spatial attention; pitch processing

Corresponding author: Adrian KC Lee, University of Washington, 1715 NE Columbia Road, Box 357988, Seattle, WA 98195, Phone: (206) 616-0102, Fax: (206) 221-6472, akclee@uw.edu.

1. 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 (Yantis et al., 2002; Petkov et al., 2004; Wu et al., 2007) 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 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 covarying 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.

2. Materials and Methods

2.1 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 *Methods*). We also excluded two subjects due to data being too noisy for reliable analysis.

2.2 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 Figure 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 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 "E"s would be three, since there was one "E" in the first three letters of the high-pitched stream, and two "E"s in the last three letters of the low-pitched stream. Note that attending to the incorrect stream during either the first or second triplet would result in an incorrect response (in this example, a zero or two).

One second after the offset of the final letter in the target and masker streams, a response circle appeared to instruct subjects to respond using a button box (ranging from 0 to 3, i.e., chance at 25%) with the number of "E"s heard (for diamond or X cues) or the number of fixation dot flashes (for the square cue). This response circle was used to temporally separate out cortical activity associated with button presses. After the response circle disappeared, the screen returned to a presentation of the fixation dot for 1400 ms before the onset of the next trial. This time period was used to allow the subjects to recover from responding to the previous trial prior to the onset of the next trial.

Target features (left- and right-direction, high- and low-pitch), attention conditions (maintain versus switch) as well as visual dot flash counting conditions were randomized across trials. Subjects performed eight 6-minute behavioral runs while in the MEG. This consisted of 352 trials total, with an equal distribution of left, right, high-pitch, and low-pitch conditions among the 256 auditory trials, plus 96 flash-counting visual trials (with 64 zero-flash and 32 nonzero-flash count trials). Note that the trials with zero fixation dot flashes constitute a fixation condition, as auditory and visual stimuli during the target and masker streams were equivalent to auditory trials, but subjects were not performing an auditory task. Prior to M-EEG scanning, subjects practiced the task, and were trained to respond only when the response circle appeared as well as maintain center fixation.

2.3 Stimulus generation

Auditory stimuli were generated using tokens from a single female talker in the ISOLET v1.3 corpus (Cole et al., 1990) with speaker and letters ("AEIJKMOQRUXY") chosen such that all tokens were as close as possible to, but no longer than, 400 ms in duration (trimmed of leading and trailing silence). All letters that rhymed with the target letter "E" were eliminated. Each letter was resampled to 24.4 kHz, monotonized, windowed with a 10 ms $\cos^2$ envelope, pitch shifted up and down using Praat software (Boersma and Weenink, 2009) to create tokens at 200 Hz and 200 Hz ± 4.25 semitones, and letters were matched in intensity (presentation level 75 dB SPL). Target and masker streams were also amplitude - modulated at either 37 or 43 Hz. This was to facilitate a complementary phase-locking analysis, the results of which are not reported here due to insufficient signal-to-noise to meaningfully compare the conditions of interest (space and pitch switch versus maintain attention, as well as space or pitch versus the "control" visual flashing condition). During spatial conditions, streams (with fundamental frequency 200 Hz) were delayed by 300 μs in the right or left channel to elicit a leftward or rightward spatial percept, respectively, without modifying the intensity delivered to each ear on any given trial. Auditory stimuli were presented using sound-isolating tubal insertion earphones (Nicolet Biomedical Instruments) with digital-to-analog conversion and amplification (Tucker-Davis Technologies). Auditory stimuli were presented against a constant Gaussian white noise background (20 dB SNR) to mask ambient noise; this noise was also inverted at one ear to generate interaural differences that caused it to "fill the head," instead of having a distinct location. Visual stimuli were presented using PsychToolbox (Brainard, 1997) back-projected onto a screen 1 m in front of subjects using a PT-D7700U-K (Panasonic) projector.

2.4 M-EEG and MRI data acquisition

During the behavioral task, MEG and EEG data were recorded simultaneously. MEG and EEG provide complementary measurements of primarily cortical activity, and combining these measures with an anatomical MRI constraint provides source localization that more precisely matches fMRI compared to using MEG or EEG alone (Sharon et al., 2007; Molins et al., 2008). M-EEG data were recorded inside a magnetically shielded room (IMEDCO) using a 306-channel dc-SQUID VectorView system (Elekta-Neuromag). Four head-position indicator (HPI) coils were mounted on a 60-channel EEG cap (Brain Products), and a 3Space Fastrak (Polhemus) was used to record locations of cardinal landmarks (nasion, left/right periauriculars), EEG electrodes, HPI coils, and at least 100 additional scalp points to coregister the M-EEG sensors with individualized structural MRIs. HPI coils were used to record the subject's head position continuously relative to the MEG sensors. In a separate session, anatomical information was obtained using three structural MRI brain scans using a 3T Philips scanner. For each subject, one standard structural multi-echo magnetization-prepared rapid gradient echo scan and two multi-echo multi-flip angle (5° and 30°) fast low-angle shot scans were acquired to provide the necessary contrast for the boundary-element models that accurately characterize MEG and EEG forward field patterns.

2.5 Data processing and analysis

The M-EEG data were analyzed using MNE (http://www.nmr.mgh.harvard.edu/mne; see Lee et al., 2012). Data were first processed using signal space separation (Taulu and Kajola, 2005) to remove environmental artifacts and translate the MEG coordinate frame for each subject to a constant frame. Data were then low-passed (55 Hz) and signal space projection was used to exclude heartbeat and blink artifacts identified using electro-cardiogram and electro-oculogram (EOG) data. Noisy M-EEG channels were excluded from subsequent processing, which should have a minimal impact on localization due to the spatial redundancy of M/EEG measurements (Nenonen et al., 2007). Trials were rejected for incorrect behavioral responses and supra-threshold EEG (150 μV) or MEG (15 pT for magnetometers, 5.0 pT/cm for gradiometers) signals. For each subject, waveforms for each condition were averaged while equalizing trial counts for each condition to avoid bias.

A cortical M-EEG source space was constructed using dipoles with 7 mm spacing, yielding roughly 3,000 dipoles per hemisphere. These were constrained to be normal to the cortical surface located along the gray/white matter boundary segmented from the structural MRI (Dale et al., 1999) using Freesurfer (http://surfer.nmr.mgh.harvard.edu/). Dipole currents in this whole-brain source space were then estimated from the preprocessed M-EEG data. To do this, we used an anatomically constrained minimum-norm linear estimation approach (Hämäläinen and Sarvas, 1989; Hämäläinen and Ilmoniemi, 1994; Dale et al., 1999) with sensor noise covariance estimated from 200 ms epochs prior each trial onset while subjects were not engaged in a behavioral task (i.e., data captured in this window reflect sensor noise and brain activity not related to the auditory attention task). Source localization data were then mapped to an average brain using a non-linear spherical morphing procedure (20 smoothing steps) that optimally aligns individual sulcal-gyral patterns (Fischl et al., 1999) and data were binned into 20-ms averaged windows.

2.6 M-EEG statistics–spatiotemporal clustering

Pairwise comparisons between spatiotemporal activations were performed using a non-parametric spatial clustering permutation test with a maximal statistic (Nichols and Holmes, 2002)extended to incorporate the temporal dimension (see Larson and Lee, 2013 for details). This method addresses multiple comparisons by conservatively controlling the family-wise error rate (FWER) while identifying regions of large spatial and consistent temporal activation. Briefly, a paired t-test across subjects contrasting maintain versus switch

activations at each point in space (cortical dipoles) and time with a "hat" relative minimal variance adjustment of 10^{-3} (Ridgway et al., 2012) created a statistical map that was thresholded at p < 0.05 (uncorrected). Putatively significant points were then clustered based on geodesic spatial (5 mm along the cortical surface) and temporal (20 ms) proximity (and t-value sign), with each cluster scored as the cumulative magnitude t-score of all included points. This procedure was repeated for 512 condition relabelings (with ten subjects and a symmetric two-tailed test, $2^{10-1} = 512$), and the distribution of maximal cluster scores from these permutations provided a multiple-comparisons corrected p-value as the proportion of maximal distribution scores greater than or equal to that of the given cluster.

A similar procedure was used to correct for multiple comparisons across time in the time-cluster (or regions of interest) time-course analyses. Specifically, a permutation-based temporal clustering test based on adjacent time samples being neighbors was used. This is equivalent to carrying out the spatiotemporal procedure outlined above with only a single "spatial location" present.

To estimate cluster location in MNI Talairach coordinates, the spatial center of mass was calculated by weighing each cluster vertex by the sum of its significant t values through time. Consistent with the Freesurfer analysis pipeline, this center of mass was calculated on the equivalent inflated spherical surface to preserve the surface-based topology underlying the analysis. As RTPJ locations reported attention-manipulating tasks vary by at least 22.7 mm (Mitchell, 2008), we defined an anatomical label for RTPJ to help establish the spatial correspondence between previous findings and our results. We established the RTPJ location by fitting the smallest possible sphere around previously reported MNI coordinates, and selecting surface vertices falling within that sphere, as in a previous study (Larson and Lee, 2013). This is only used to relate results from the whole-brain statistical approach to previous studies that have implicated RTPJ involvement in attention.

In addition to this whole-brain statistical approach, we also made use of a *post-hoc* region-of-interest (ROI)-based statistical approach. This was done to examine activation in two areas from our previous work (Lee et al., 2013), namely the superior temporal sulcus (STS) and superior precental sulcus region (including frontal eye fields). We built these regions of interest from the standard anatomical parcellations in the Freesurfer software, analyzing activation bilaterally for completeness.

To evaluate and display the time courses of identified significant spatiotemporal clusters and ROIs, trials were epoched (baseline-corrected), averaged across trials (analysis time-locked to trial onset), and projected onto the cortical surface, and the magnitude of the currents from each of the vertices in the cluster (significant at any point in time) were averaged. To compensate for differences in overall SNR across subjects, the resulting traces were normalized (divided by the mean value across time for each subject) before being averaged together for display, yielding arbitrary units (AU) for the axes.

3. Results

3.1 Behavioral task

M/EEG and behavioral data from ten subjects were analyzed (Figure 2). Results from a two-way repeated-measures ANOVA show there was no significant difference between the space and pitch conditions ($p = 0.605$), or interaction between these two factors ($p = 0.118$). While there was an apparent trend of lower performance in switching trials (see mostly negative y-axis values in Figure 4), there was also no significant difference between switching versus maintaining attention ($p = 0.097$). Thus, this suggests that the task was approximately

equally difficult for subjects in the maintain- and switch-attention conditions, and in the attend-space and attend-pitch conditions.

3.2 Cortical dynamics of switching attention

Contrasting the maintain- versus switch-attention conditions when subjects attended based on spatial information (stimulus ITD), we found that RTPJ ($p = 0.0195$ FWER corrected, center of mass x = 47.0, y = –49.2, z = 22.7) was more active when subjects switched attention than when they maintained attention (Figure 3). Moreover, we found that the normalized difference in behavior ([switch – maintain]/maintain) was significantly correlated with the normalized difference in neural activation in RTPJ during the switch period ($p < 0.02$; Figure 4), matching closely with the results found in our previous study on switching spatial attention (Larson and Lee, 2013). When contrasting maintain- versus switch-attention on trials where subjects attended based on the pitch of the streams, we found that the left inferior parietal supramarginal part (LIPSP; slightly dorsal of Wernicke's area) was more involved when maintaining attention to sound based on pitch ($p = 0.0352$ FWER corrected, center of mass x = –54.0, y= –40.2, z = 32.1). The activation of this cluster did not significantly correlate with normalized behavior performance on the pitch task ($p > 0.5$). We observed no significant clusters other than RTPJ and LIPSP in our whole-brain contrasts of switch- versus maintain-attention for either space or pitch.

We then performed an analysis comparing the time courses of each significant cluster in each condition to examine the range of times that each cluster was significantly active over the entire task duration when different features were available (spatial and pitch stimuli). We found that the RTPJ cluster is significantly more active on spatial trials in the switch attention condition from 2.54 to 3.98 s ($p < 0.002$, FWER corrected), unsurprisingly including the switch period (2.75–3.35 s) that was used to generate the cluster, whereas the LIPSP cluster is not significant during this period ($p > 0.05$, FWER corrected). When contrasting the maintain- and switch-attention conditions when attending based on pitch, LIPSP is significantly more active during the maintain- versus switch-attention task from 1.84 to 3.94 s ($p < 0.002$ FWER corrected, again including the 2.75–3.35 s switch period), whereas RTPJ activation is not significantly different during this period ($p > 0.05$, FWER corrected). We then performed ROI-based analyses of the activation time courses in the superior temporal sulcus (STS), and the superior precentral sulcus (as a proxy for FEF) in each hemisphere, but did not find significant differences between maintaining and switching attention in the spatial or pitch conditions ($p > 0.05$ all). We also performed a *post-hoc* whole-brain contrast of switch-space and switch-pitch conditions, but this measure (which has lower sensitivity than time course-based analyses) did not yield any significant clusters at the $p < 0.05$ FWER-controlled threshold.

4. Discussion

The cortical dynamics underlying the switching of auditory attention is not yet well understood. In particular, the generality of attention switching mechanisms across different stimulus features and/or across different sensory modalities is unclear. Here we offer evidence that auditory attention switching involves separate structures depending on whether stimuli differ based on spatial or pitch information, namely RTPJ and LIPSP. Interestingly, we found that these regions had a double dissociation, where RTPJ was significantly more active on trials where subjects switched attention when spatial features were available but not when pitch features were available, and vice-versa for LIPSP. This offers evidence that top-down selective attention mechanisms operate differently based on the type of information that is available, even within a single modality.

In our task, we found a correlation between RTPJ activation and behavioral performance, consistent with our previous study (Larson and Lee, 2013). Our prior study used a substantially different paradigm, where listeners were cued to switch attention between hemifields during a preparatory period prior to the target and masker auditory stimuli (single digits) being presented. This previous study motivated the temporal separation of the switch attention cue (at beginning of the trial) from the switch attention period (over a second later) in the current paradigm, as it isolated the neural responses due to visual cueing from the neural responses involved in endogenous attention switching. The agreement found between these two studies suggests that RTPJ plays a role in the switching of auditory spatial attention, independent of the multiple differences between these tasks. This finding is consistent with evidence from vision, where RTPJ operates as a "filter" for incoming stimuli by suppressing attention switching to distractors (Shulman et al., 2007) and acts as a "circuit breaker" redirecting attention to relevant stimuli outside the current focus of attention (Corbetta and Shulman, 2002). Notably, in visual spatial attention studies RTPJ is triggered by bottom-up salient stimuli as opposed to endogenous attention control (e.g., during visual search). Thus endogenous auditory spatial attention tasks appear to engage a supra-modal spatial attention network by effecting attention changes in a similar manner to bottom-up visual stimuli. This finding is consistent with other studies in audition showing that TPJ mediates endogenous attention shifts (Salmi et al., 2009), with bilateral TPJ active when reorienting attention after an invalid cue (Mayer et al., 2009), as well as another study showing activation in the intraparietal sulcus (possibly including RTPJ) during both endogenously and exogenously cued auditory attention shifts (Huang et al., 2012).

Although there was a double dissociation between activation of RTPJ and LIPSP in the task as a whole, the temporal activation patterns of these regions suggest that they may have played different functional roles. RTPJ became significantly more active in the switch-attention condition only 200 ms prior to the switch gap period on space trials (about a second after the target and masker streams began), whereas LIPSP became more active in the maintain-attention condition during presentation of the first target letter (about a second before the switch period). While we hypothesized that different mechanisms may be involved in switching attention based on spatial and non-spatial features, we did not have an *a priori* hypothesis regarding how switching attention based on pitch would operate. However, our results suggest that different behavioral and/or neural strategies may have been employed by subjects for these attention tasks. Specifically, in the spatial condition, RTPJ likely acts as a "circuit breaker" to trigger the spatial reorientation of the focus of attention to the new location and stimulus of interest. However, in the pitch condition, subjects may use speech processing or episodic memory centers to maintain a representation of the target speaker pitch when they are required to maintain attention. The significant LIPSP activation begins at the onset of the first target letter (when selective attention becomes challenging) and persists through the gap period, and our observed LIPSP cluster overlaps with left-lateralized structures that are involved in speech processing (e.g., Zheng et al., 2013)and episodic memory retrieval (e.g., Hutchinson et al., 2009). In this way, switching attention based on pitch (and potentially other non-spatial) features may involve distinct neural processing and circuitry from spatial attention. Pitch-based attention may also nonetheless engage a shared supramodal attention control network, but our finding that RTPJ was not differentially involved during the pitch task suggests that such engagement occurs elsewhere in the attention network. It could be useful in future studies to test the extent to which our findings may be affected by episodic memory or speech-related processes (e.g., by using non-speech stimuli).

Multiple neuroimaging studies involving judging spatial location or pitch have examined how these spatial and pitch features are neurally encoded. When judging sequences with location or pitch changes, primary auditory areas such as planum temporale and Heschl's

gyrus are more active (Warren and Griffiths, 2003). Localizing compared to recognizing stimuli evokes different response levels in premotor areas (Maeder et al., 2001). Premotor areas are also more active when attending to stimuli based on space versus pitch differences, with bilateral superior temporal areas active in both tasks (Degerman et al., 2006). These studies thus suggest that partially distinct regions participate in processing location and pitch information. However, these studies contrasted activity during epochs in which both the physical stimuli and the top-down task demands differed. The observed differences could thus be attributed to changes in bottom-up processing of stimuli or in top-down task demands, making it difficult to compare these results to the present study.

Other studies have identified neural substrates involved in selective attention by keeping the presented stimuli constant while changing the top-down task demands. When auditory attention is based on either the location or the pitch of tone sequences, right premotor areas are preferentially engaged (Zatorre et al., 1999). During adaptation to repeated presentation of stimuli, there is evidence for "what" and "where" selection pathways that show enhanced tuning to spatial and pitch features in posterior and anterior STS, respectively (Ahveninen et al., 2006). However, these studies do not clearly separate activity evoked when a listener prepares to attend based on a specific feature from the activity evoked by auditory object selection occurring while the stimuli are actually presented, which could result in multiple differences in the related neural processing reported. Nonetheless, these studies indicate a partially separable anterior–posterior or dorsal–ventral what–where split in auditory stream processing (Rauschecker and Tian, 2000; Rauschecker, 2011).

In our study, the pitch condition provided listeners with salient cues for auditory stream segregation but no spatial information, and thus could be considered to provide predominantly "what" information. Here we also found a what-where split, but the difference occurred across hemispheres (left primarily for pitch, right primarily for space), a notable departure from these other studies. That may be because, in the present study, we focused on preparatory attention control signals, that is, during time epochs during which neural activity is driven by endogenous task demands instead of bottom-up sensory input. This helps to isolate the control signals responsible for switching attention between competing auditory stimuli based on (upcoming) stimulus features. We speculate that the resulting hemispheric asymmetry is driven by differences in the underlying neural strategy deployed in the tasks – namely, gating spatial attention via RTPJ or maintaining a working memory representation using LIPSP. It should also be noted that, as with any neuroimaging paradigm, regions other than the ones reported here (RTPJ and LIPSP) also likely play an important role in these tasks. Specifically, the conservatively multiple-comparisons corrected whole-brain analysis used here may have excluded additional regions that are involved. For example, our whole-brain direct contrast of switch-space and switch-pitch attention conditions did not reveal the same double dissociation identified by analyzing cluster time courses (like in a traditional ROI-based analysis)—this is likely due to the substantially lower sensitivity of the whole-brain approach compared to the ROI analysis as opposed to a true lack of difference between these two tasks. Additional future tasks or analyses (e.g., time–frequency or functional connectivity methods) could shed additional light on these issues.

In extensive work done primarily in vision, it has been shown that a dorsal cortical network mediates attention reorientation (Corbetta et al., 2008) based on endogenous, top-down goals. Recent studies comparing audition and vision indicate that cortical regions involved in attending during auditory and visual and tasks are mostly conserved (Shomstein and Yantis, 2006; Wu et al., 2007), suggesting that attentional control involves a network that operates supramodally. However, our results allow us to speculate that the extent to which supramodal mechanisms of attentional control are recruited may depend on the information

available in the auditory stimuli. Specifically, the typical concordance between auditory and visual spatial information would make it natural for auditory spatial attention to recruit or share cortical mechanisms involved in processing spatial information that are also used in vision. However, pitch information has no immediate visual correlate in natural stimuli, and thus it would be reasonable for a structure or set of structures unique to the auditory system to mediate the flow of pitch-based information. Evidence has been shown before for differential "what" and "where" auditory stimulus processing (Ahveninen et al., 2006). However, the manner in which the flow and use of what and where information is controlled by potentially different mechanisms based on top-down task goals remains an open question. Here, we observed a double dissociation between the cortical structures that differentially responded to tasks requiring switching versus maintaining attention based on spatial (RTPJ) and pitch (LIPSP) features. These results add some support for the hypotheses that top-down attention mechanisms differ based on auditory stimulus properties available for segregation, and that auditory attention involves similar regions to visual attention processing when spatial information is available.

Acknowledgments

Abbreviations

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

References

Ahveninen J, Jaaskelainen I, Raij T, Bonmassar G, Devore S, Hämäläinen M, Levanen S, Lin F, Sams M, Shinn-Cunningham B, Witzel T, Belliveau J. Task-modulated "what" and "where" pathways in human auditory cortex. Proceedings of the National Academy of Sciences of the United States of America. 2006; 103:14608–13. [PubMed: 16983092]

Boersma P, Weenink D. Praat: doing phonetics by computer (Version 5.1.13). 2009:x.

Brainard DH. The Psychophysics Toolbox. Spatial Vision. 1997; 10:433–6. [PubMed: 9176952]

Cole, R.; Muthusamy, Y.; Fanty, M. The ISOLET spoken letter database. Oregon Graduate Institute of Science and Technology; 1990.

Corbetta M, Patel G, Shulman GL. The reorienting system of the human brain: from environment to theory of mind. Neuron. 2008; 58:306–24. [PubMed: 18466742]

Corbetta M, Shulman GL. Control of goal-directed and stimulus-driven attention in the brain. Nature Reviews Neuroscience. 2002; 3:201–15.

Dale A, Fischl B, Sereno M. Cortical surface-based analysis. I. Segmentation and surface reconstruction. Neuroimage. 1999; 9:179–194. [PubMed: 9931268]

Degerman A, Rinne T, Salmi J, Salonen O, Alho K. Selective attention to sound location or pitch studied with fMRI. Brain Res. 2006; 1077:123–34. [PubMed: 16515772]

Driver J, Spence C. Attention and the crossmodal construction of space. Trends in Cognitive Sciences. 1998; 2:254–262. [PubMed: 21244924]

Hämäläinen M, Sarvas J. Realistic conductivity geometry model of the human head for interpretation of neuromagnetic data. IEEE Transactions in Biomedical Engineering. 1989; 36:165–171.

Hämäläinen MS, Ilmoniemi RJ. Interpreting magnetic fields of the brain: minimum norm estimates. Med Biol Eng Comput. 1994; 32:35–42. [PubMed: 8182960]

Hill KT, Miller LM. Auditory Attentional Control and Selection during Cocktail Party Listening. Cereb Cortex. 2010; 20:583–590. [PubMed: 19574393]

Huang S, Belliveau JW, Tengshe C, Ahveninen J. Brain Networks of Novelty-Driven Involuntary and Cued Voluntary Auditory Attention Shifting. PLoS ONE. 2012; 7:e44062. [PubMed: 22937153]

Hutchinson JB, Uncapher MR, Wagner AD. Posterior parietal cortex and episodic retrieval: Convergent and divergent effects of attention and memory. Learn Mem. 2009; 16:343–356. [PubMed: 19470649]

Kiesel A, Steinhauser M, Wendt M, Falkenstein M, Jost K, Philipp AM, Koch I. Control and interference in task switching--a review. Psychol Bull. 2010; 136:849–874. [PubMed: 20804238]

Koch I, Lawo V, Fels J, Vorländer M. Switching in the cocktail party: exploring intentional control of auditory selective attention. J Exp Psychol Hum Percept Perform. 2011; 37:1140–1147. [PubMed: 21553997]

Larson E, Lee AKC. The cortical dynamics underlying effective switching of auditory spatial attention. Neuroimage. 2013; 64:365–370. [PubMed: 22974974]

Lee AK, Larson E, Maddox RK. Mapping cortical dynamics using simultaneous MEG/EEG and anatomically-constrained minimum-norm estimates: an auditory attention example. J Vis Exp. 2012:e4262. [PubMed: 23128363]

Lee AKC, Rajaram S, Xia J, Bharadwaj H, Larson E, Shinn-Cunningham BG. Auditory selective attention reveals preparatory activity in different cortical regions for selection based on source location and source pitch. Front Neurosci. 2013; 6:190. [PubMed: 23335874]

Maeder PP, Meuli RA, Adriani M, Bellmann A, Fornari E, Thiran JP, Pittet A, Clarke S. Distinct pathways involved in sound recognition and localization: a human fMRI study. Neuroimage. 2001; 14:802–16. [PubMed: 11554799]

Mayer AR, Franco AR, Harrington DL. Neuronal modulation of auditory attention by informative and uninformative spatial cues. Hum Brain Mapp. 2009; 30:1652–1666. [PubMed: 18661505]

Meiran N, Chorev Z, Sapir A. Component Processes in Task Switching. Cognitive Psychology. 2000; 41:211–253. [PubMed: 11032657]

Mesgarani N, Chang EF. Selective cortical representation of attended speaker in multi-talker speech perception. Nature. 2012; 485:233–236. [PubMed: 22522927]

Mitchell JP. Activity in right temporo-parietal junction is not selective for theory-of-mind. Cereb Cortex. 2008; 18:262–71. [PubMed: 17551089]

Molins A, Stufflebeam SM, Brown EN, Hämäläinen MS. Quantification of the benefit from integrating MEG and EEG data in minimum l2-norm estimation. Neuroimage. 2008; 42:1069–77. [PubMed: 18602485]

Nenonen J, Taulu S, Kajola M, Ahonen A. Total information extracted from MEG measurements. International Congress Series. 2007; 1300:245–248.

Nichols TE, Holmes AP. Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum Brain Mapp. 2002; 15:1–25. [PubMed: 11747097]

Petkov C, Kang X, Alho K, Bertrand O, Yund E, Woods D. Attentional modulation of human auditory cortex. Nature Neuroscience. 2004; 7:658–663.

Posner MI, Snyder CR, Davidson BJ. Attention and the detection of signals. J Exp Psychol. 1980; 109:160–174. [PubMed: 7381367]

Rauschecker JP. An expanded role for the dorsal auditory pathway in sensorimotor control and integration. Hearing Research. 2011; 271:16–25. [PubMed: 20850511]

Rauschecker JP, Tian B. Mechanisms and streams for processing of "what" and "where" in auditory cortex. Proceedings of the National Academy of Sciences. 2000; 97:11800–11806.

Ridgway GR, Litvak V, Flandin G, Friston KJ, Penny WD. The problem of low variance voxels in statistical parametric mapping; a new hat avoids a "haircut. Neuro Image. 2012; 59:2131–2141. [PubMed: 22037420]

Salmi J, Rinne T, Koistinen S, Salonen O, Alho K. Brain networks of bottom-up triggered and top-down controlled shifting of auditory attention. Brain Res. 2009; 1286:155–164. [PubMed: 19577551]

Sharon D, Hämäläinen MS, Tootell RBH, Halgren E, Belliveau JW. The advantage of combining MEG and EEG: comparison to fMRI in focally stimulated visual cortex. Neuroimage. 2007; 36:1225–35. [PubMed: 17532230]

Shinn-Cunningham BG. Object-based auditory and visual attention. Trends in Cognitive Sciences. 2008; 12:182–186. [PubMed: 18396091]

Shomstein S, Yantis S. Parietal cortex mediates voluntary control of spatial and nonspatial auditory attention. J Neurosci. 2006; 26:435–9. [PubMed: 16407540]

Shulman GL, Astafiev SV, McAvoy MP, d'Avossa G, Corbetta M. Right TPJ deactivation during visual search: functional significance and support for a filter hypothesis. Cereb Cortex. 2007; 17:2625–33. [PubMed: 17264254]

Taulu S, Kajola M. Presentation of electromagnetic multichannel data: The signal space separation method. Journal of Applied Physics. 2005; 97:124905, 124905–10.

Voisin J, Bidet-Caulet A, Bertrand O, Fonlupt P. Listening in Silence Activates Auditory Areas: A Functional Magnetic Resonance Imaging Study. J Neurosci. 2006; 26:273–278. [PubMed: 16399697]

Warren J, Griffiths T. Distinct mechanisms for processing spatial sequences and pitch sequences in the human auditory brain. Journal of Neuroscience. 2003; 23:5799–5804. [PubMed: 12843284]

Wu C-T, Weissman DH, Roberts KC, Woldorff MG. The neural circuitry underlying the executive control of auditory spatial attention. Brain Research. 2007; 1134:187–98. [PubMed: 17204249]

Yantis S, Schwarzbach J, Serences JT, Carlson RL, Steinmetz MA, Pekar JJ, Courtney SM. Transient neural activity in human parietal cortex during spatial attention shifts. Nat Neurosci. 2002; 5:995–1002. [PubMed: 12219097]

Zatorre RJ, Mondor TA, Evans AC. Auditory attention to space and frequency activates similar cerebral systems. Neuroimage. 1999; 10:544–554. [PubMed: 10547331]

Zheng ZZ, Vicente-Grabovetsky A, MacDonald EN, Munhall KG, Cusack R, Johnsrude IS. Multivoxel Patterns Reveal Functionally Differentiated Networks Underlying Auditory Feedback Processing of Speech. J Neurosci. 2013; 33:4339–4348. [PubMed: 23467350]

Highlights

- Tested attention switching in audition based on spatial location and pitch

- Right temporoparietal junction is only involved in switching based on location

- Left inferior parietal supramarginal part is only engaged when subjects use pitch

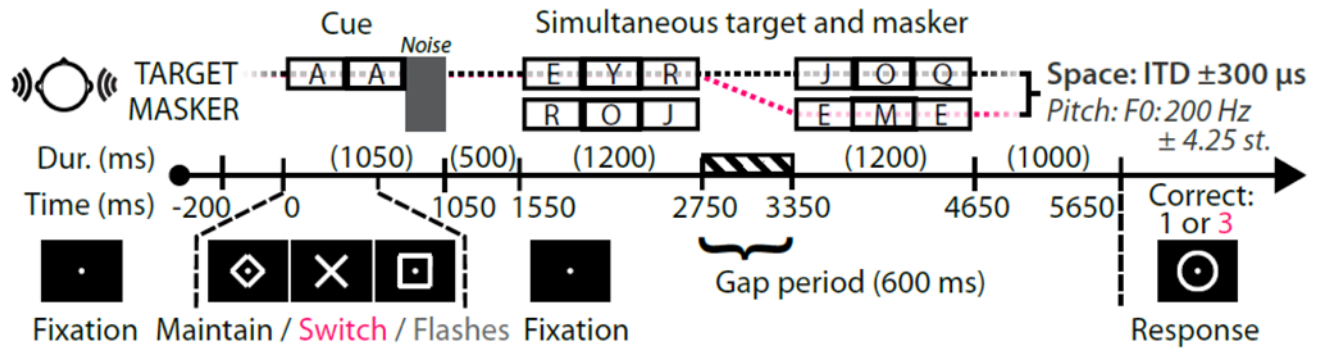- Double dissociation and neural timing suggest distinct mechanisms and/or strategies

Figure 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.
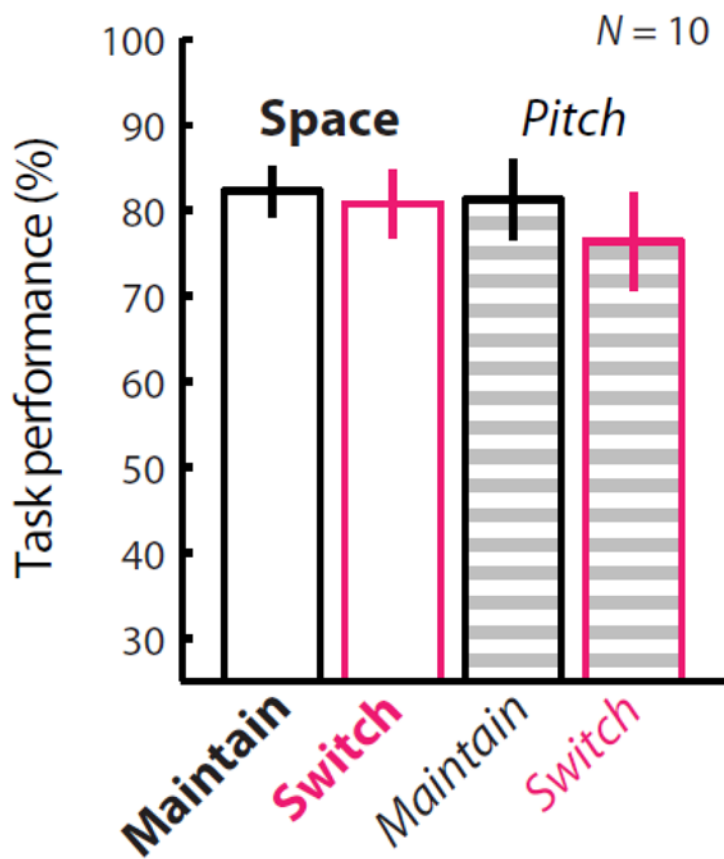
Figure 2. Attending based on space and pitch features yielded similar behavioral performance
Behavioral scores for $N = 10$ subjects show no significant differences between space or pitch trials (ANOVA $p = 0.605$) or between maintain- and switch-attention conditions ($p = 0.097$). This suggests that differences in neural activity are not likely due to differences in task difficulty. Only correct trials were included in subsequent analyses.
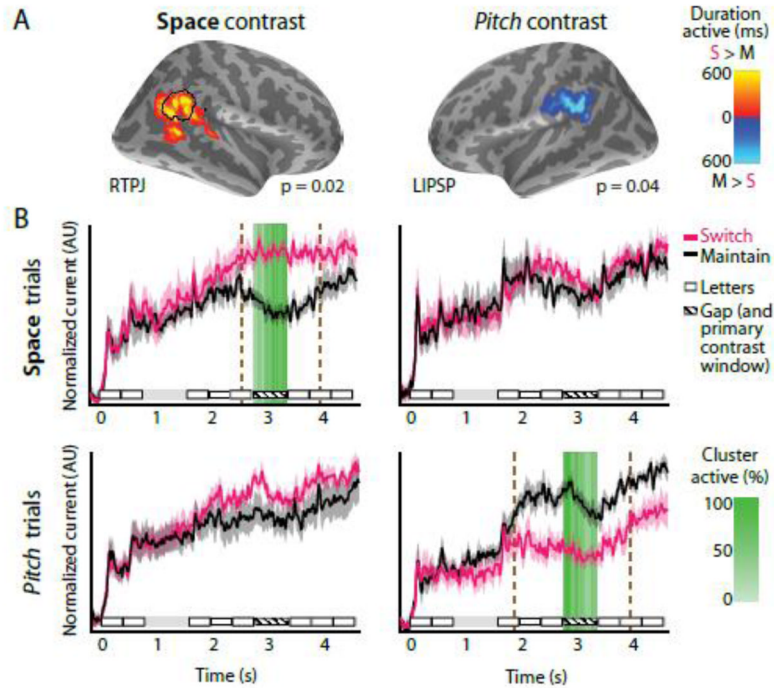
Figure 3. Significant activation differences in a whole-brain analysis of maintain- versus switch-attention conditions in conditions reveal a double dissociation when subjects attended based on space and pitch

A: RTPJ is more active when switching attention based on spatial information, while LIPSP is more active when maintaining attention in based on pitch information. The duration each vertex in the cluster was significantly active is shown by the colorbar. *B:* The time courses of neural activation for maintain (black) and switch (pink) attention. The primary contrast was performed in the gap period (hatched line), during which RTPJ and LIPSP clusters were significantly more active (switch/maintain attention contrast) only during the space and pitch conditions, respectively. The anatomical RTPJ definition (see *Methods*) is shown by a black outline. The percentage of clusters vertices significantly different as a function of time are shown using the green colorbar. A secondary temporal contrast using the activity from these clusters showed that the RTPJ cluster was significantly more active in the switch-attention condition between 2.54 – 3.98 seconds on space trials ($p < 0.002$ FWER corrected, brown dashed lines), but did not significantly differ in the pitch case. Conversely, LIPSP was significantly more active in the maintain-attention condition between 1.84 – 3.94 seconds on pitch trials ($p < 0.002$ corrected, brown dashed lines), but not on space trials.
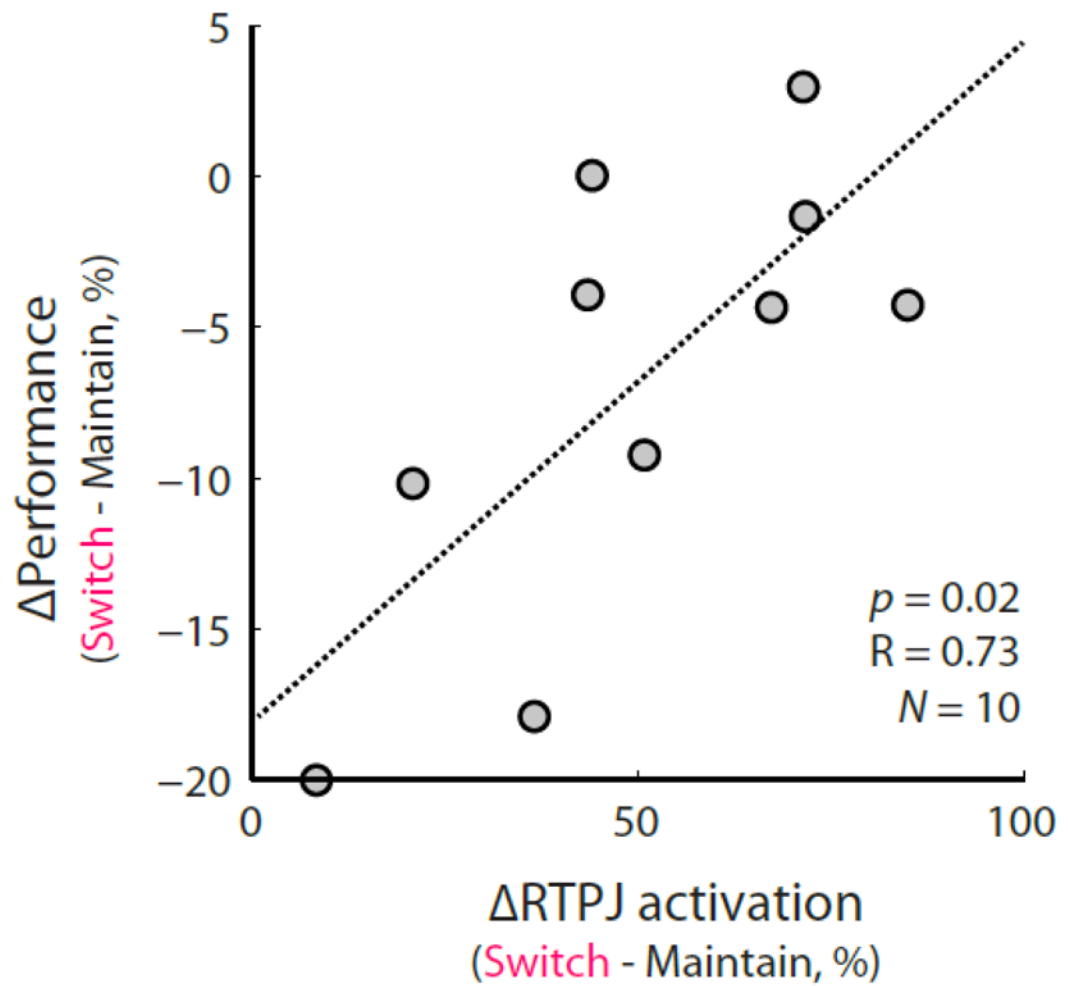
Figure 4.
The normalized behavioral performance (100 * [switch – maintain]/maintain) on space trials was significantly correlated ($p < 0.02$, R = 0.73) with the normalized difference in activity in RTPJ measured during the gap period during which subjects could switch attention.