Brain activity is related to individual differences in the number of items stored in auditory short-term memory for pitch: Evidence from magnetoencephalography

Stephan Grimault\textsuperscript{a,b,c,*}, Sophie Nolden\textsuperscript{a,b}, Christine Lefebvre\textsuperscript{a,b,f}, François Vachon\textsuperscript{a,b,d}, Krista Hyde\textsuperscript{b,e}, Isabelle Peretz\textsuperscript{a,b}, Robert Zatorre\textsuperscript{b,c}, Nicolas Robitaille\textsuperscript{a,b}, Pierre Jolicœur\textsuperscript{a,b,f}

\textsuperscript{a} CERNEC, Université de Montréal, Québec, Canada
\textsuperscript{b} BRAMS, International Laboratory for Brain Music and Sounds, Université de Montréal, Québec, Canada
\textsuperscript{c} Centre National de la Recherche Scientifique (CNRS), France
\textsuperscript{d} Ecole de psychologie, Université Laval, Québec, Canada
\textsuperscript{e} Montreal Neurological Institute, McGill University, Québec, Canada
\textsuperscript{f} Centre de recherche de l'institut universitaire de gériatrie de Montréal (CRIUGM), Québec, Canada

\textbf{A R T I C L E  I N F O}

Article history:
Accepted 8 March 2014
Available online 15 March 2014

Keywords:
Auditory
Short-term memory
Working memory
Pitch
MEG

\textbf{A B S T R A C T}

We used magnetoencephalography (MEG) to examine brain activity related to the maintenance of non-verbal pitch information in auditory short-term memory (ASTM). We focused on brain activity that increased with the number of items effectively held in memory by the participants during the retention interval of an auditory memory task. We used very simple acoustic materials (i.e., pure tones that varied in pitch) that minimized activation from non-ASTM related systems. MEG revealed neural activity in frontal, temporal, and parietal cortices that increased with a greater number of items effectively held in memory by the participants during the maintenance of pitch representations in ASTM. The present results reinforce the functional role of frontal and temporal cortices in the retention of pitch information in ASTM. This is the first MEG study to provide both fine spatial localization and temporal resolution on the neural mechanisms of non-verbal ASTM for pitch in relation to individual differences in the capacity of ASTM. This research contributes to a comprehensive understanding of the mechanisms mediating the representation and maintenance of basic non-verbal auditory features in the human brain.

\textcopyright 2014 Elsevier Inc. All rights reserved.

\textbf{Introduction}

Short-term memory (STM) is a central cognitive function that provides an interface between perception, action, and long-term memory. Auditory STM (ASTM) in particular is an important building block for any function that requires the integration of sounds over time, such as the comprehension of spoken sentences, learning a new language, or the appreciation of music. Various models of STM have been proposed (for a review, see D’Esposito, 2007). Early models of STM proposed that specialized subsystems maintain information from different sensory inputs (e.g., Brooks, 1968) and that a central control system coordinates the flow of information through the system (Atkinson & Shiffrin, 1968). Baddeley (1986, 2001) and Baddeley and Hitch (1974) proposed that there are three main components involved in STM: 1) a “phonological loop,” that involves an auditory memory store and articulatory rehearsal process, 2) a “visuospatial sketchpad,” that processes visual–spatial information and, 3) a “central executive,” that manages and manipulates the information. However, Baddeley’s model has several limitations. For example, this model lacks detail with regard to the retention of nonverbal acoustic material, and it may require an ever-expanding number of specialized buffers for different types of information.

More recent models of STM (Cowan, 2008; D’Esposito, 2007; D’Esposito et al., 2000; Goldman-Rakic, 1987; Petrides, 1991, 2005; Postle, 2006; Postle et al., 1999; Ruchkin et al., 2003) do not rely on separate storage buffers, but rather propose that active maintenance or storage of task-relevant representations occurs in the same brain areas that support the perceptual representation of information. This type of neural system would be more flexible and efficient relative to a system that transfers information back and forth between dedicated storage buffers. Thus, the short-term retention of basic acoustic dimensions, such as pitch, would be reflected, in part, by the prolonged activation of the brain regions that support the processing of pitch. This would be distinct from those processes involved in the memory for attributes of objects perceived through other senses (e.g., for color).

The main goal of the present study was to localize the functional neural correlates of basic non-verbal ASTM for pitch. Most previous
brain imaging studies have explored ASTM in the context of speech, while relatively fewer studies have specifically investigated non-verbal ASTM for pitch. For example, Zatorre et al. (1994) used positron emission tomography (PET) to investigate the neural correlates of ASTM in the context of melodic perception and memory for pitch. In one task, the pitch of two consecutive notes was compared, and in another task, the pitch of the first and last notes of a six-note sequence was compared. The consecutive-note condition resulted in right frontal lobe activation, whereas the first versus last-note condition implicated various regions including the right frontal and temporal cortices as well as parietal and insular cortices. Zatorre et al. (1994) concluded that pitch comparisons involve a neural network that includes the right prefrontal cortex, but that the active retention of pitch involves the interaction of right temporal and frontal cortices. More recently, Gaab et al. (2003) investigated melodic pitch memory using functional MRI (fMRI). Participants were scanned while they listened to short melodies in which they had to judge whether the last or second to last tone was the same or different as the first tone. As in the Zatorre et al. (1994) study, Gaab et al. (2003) found activity in superior temporal, frontal, and parietal cortices, in response to their pitch memory task. The last authors also found activity in the supramarginal gyrus and the cerebellum. The pitch memory task used in Gaab et al.’s (2003) study above relied more on the left hemisphere versus right hemisphere regions. In an fMRI study, Koelsch et al. (2009) investigated the functional architecture of ASTM for both non-verbal and verbal auditory information during rehearsal and articulatory suppression. Participants were presented with strings of four sung syllables and were asked to recall either the pitches (‘non-verbal’) or syllables (‘verbal’). The rehearsal of both verbal and non-verbal materials implicated a brain network including the premotor cortex, the planum temporale, the inferior parietal lobe, the anterior insula, subcortical structures, and the cerebellum. Overall, the results suggested that the rehearsal and storage of both verbal and non-verbal information rely on largely overlapping neuronal networks, in particular in sensorimotor-related brain areas. The Koelsch et al. (2009) study showed mostly bilateral activations in all tasks. Although Koelsch et al. (2009) used task instructions that asked for recall of only pitch or syllables, it is not entirely clear how the experimental procedure would preclude retention of all attributes in all tasks, which complicates the interpretation of the results. Taken together, the above studies point towards a network of brain regions involved in non-verbal ASTM for pitch, particularly in frontal and temporal cortices. However, none of the previous studies reviewed above investigated which brain areas might respond to parametrically increasing memory load or relate to individual differences in the capacity of ASTM. This is the main objective of the present study. We recently conducted a series of event-related potential (ERP) studies in our laboratory to investigate the time-course of neural activity on an ASTM task in which the memory load was parametrically varied (Guimond et al., 2011; Lefebvre et al., 2013; Nolden et al., 2013). Recently, our approach, described in Grimault et al. (2009a) was also adopted by Albouy et al. (2013). Our experimental protocol was based on three key principles allowing us to focus on the maintenance of objects in ASTM: 1) to isolate brain activity that varies with memory load while keeping other aspects of processing constant; 2) to focus on brain activity during the period of retention in a memory task; and 3) to use stimuli that would engage minimally cognitive or perceptual systems other than those directly involved in ASTM. The ERP results revealed fronto-central brain activity reflected in a sustained anterior negativity during the retention phase of the ASTM task that increased in amplitude (became more negative) with increasing memory load. These ERP studies demonstrated the viability of our ASTM paradigm, and provided valuable information on the time course of functional brain activity related to ASTM. However, the ERP technique did not allow us to localize precisely where in the brain this activity takes place.

In the present study, we used magnetoencephalography (MEG) to localize the brain areas involved in the retention phase of non-verbal ASTM for pitch. MEG provides a much finer spatial resolution relative to the ERP technique, and also has a much finer time resolution relative to other imaging techniques such as functional MRI. Thus, we were able to examine the time course of ASTM-related brain activity in concert with brain localization. A similar approach using MEG was shown to be very effective in studies of visual-STEM conducted in our laboratory (Grimault et al., 2009a; Robitaille et al., 2009, 2010). To our knowledge, the present study is the first to use MEG source-localization analysis to investigate the neural correlates of non-verbal ASTM of pitch as a function of parametrically increasing memory load. Based on previous brain imaging findings on non-verbal ASTM (e.g. Guimond et al., 2011; Koelsch et al., 2009), we expect to find brain activity in frontal and temporal cortices that would correlate with performance in an ASTM task in which memory load was parametrically varied. Importantly, we also examined how brain activity was related to individual differences in ASTM ability, which allowed us to establish a strong link between observed brain activity and its functional significance for maintenance in ASTM.

Methods

Participants

Sixteen university students (6 females and 10 males) participated in the present study. All reported to have normal hearing. Data for three participants were excluded because more than 50% of their trials were rejected due to ocular artifacts (horizontal movements and/or blinks), or excessive head motion. This left a total of 13 participants (5 females, and 8 males) with a mean age of 23 years (SD = 2.6 years). All participants were right-handed. All participants gave written informed consent for procedures that were vetted by the appropriate ethics committee.

Stimuli and ASTM task

Participants were presented with two sound sequences separated by a 2000 ms silent retention period. Each sound was 100 ms in duration (including 10 ms rise and fall times) and adjacent sounds were separated by 100 ms of silence. The sounds were pure tones with frequencies selected so as to be ‘non-musical’ in the sense that intervals between the tones did not correspond to the intervals in the standard Western 12-tone-equal-tempered scale. We determined tone frequencies by dividing a two octave range into seven equal logarithmic steps per octave as follows: 380, 419, 463, 511, 564, 623, 688, 759, 838, 925, 1022, 1128, 1245, and 1375 Hz. The use of these stimuli minimized the potential recruitment of memory structures specialized for verbal or phonological information, as well as the potential contributions from memory for musical structures associated with years of listening to Western music. This approach enabled a better focus on memory for basic attributes of sound. Stimuli were presented at the most comfortable level for each participant.

On each trial, two sound sequences were presented to each participant, and the task was to decide whether the sequences were the same or different (see Fig. 1). The first sound sequence (“memory sequence”) consisted of seven sounds, the first of which was always white noise. The remaining six sounds were either all white noise, or included two or four pure tones among the white noise segments. Half of the trials contained “interspersed” sequences, in which the pure tones could appear at any position after the first white noise and in which tones could be separated by one or more white noise bursts (for an example of an “interspersed” sequence, see the sequence for Load 4 in Fig. 1). In the other half of the trials, the “contiguous” trials, the pure tones appeared without interspersed white noises at the last positions of the sequence (for an example of a contiguous sequence, see the sequence for Load 2 in Fig. 1). The second sequence (“test sequence”) consisted of only pure tones (no white noise) from the first sound sequence, that were either presented in the same or different orders relative to the first sound sequence. In the case of differences in the first
sound sequence, S1, and the second sequence, S2, the order of two contiguos sounds in S1 was inverted to create S2. For an example of "same" sequences, see condition Load 2 in Fig. 1, and for an example of "different" trials, see condition Load 4 in Fig. 1. Memory load was manipulated by varying the number of pure tones presented in the first sequence (0, 2, or 4 pure tones). Following each trial, participants responded using a button press whether the sequences were the "same" or "different." When the first sequence was comprised of all white noise, the test sequence contained a single pure tone and the correct response in this case was always "different." For each subject, the average number of items held in memory was measured using Cowan's (2001) 'K' coefficient based on their accuracy in the memory task. 'K' refers to the estimated number of objects encoded into memory and is calculated according to the formula: $K = \text{load} \times (\text{Hits} - \text{False Alarm})$. "Hits" were the proportion of 'different' trials on which participants correctly detected a change, and "False Alarms" were the proportion of 'same' trials on which the participant incorrectly reported a change. The addition of the white-noise sounds allowed us to create sound sequences that had the same number of sound events and the same total duration in all load conditions, but that nevertheless varied in load for the sound objects to be remembered. This method has been used successfully to isolate load-related brain activity related to ASTM revealed the task at a sampling rate of 1200 Hz using a low-pass filter at 300 Hz. Vertical and horizontal electro-occulograms were also recorded (HEOG, VEOG, using two pairs of bipolar electrodes). Third-order gradient noise reduction (computed with CTF software) was applied to all MEG signals. These signals were next segmented into individual trials from $-1500$ to $+3700$ ms (relative to the onset of the retention interval), and 'baseline-corrected' by subtracting the mean activity at 200 ms preceding the first sound of the encoding sequence, from the signal of each trial (done separately for each sensor). Following careful visual inspection of the VEOG and HEOG, trials with movement artifacts (e.g., eye blinks, eye movements, head movements) were excluded. On average, 32% of the data were lost as a result of these exclusion steps. We anticipated this fairly high rate of exclusion because of the long time period during which participants needed to keep from blinking, and we compensated by performing a larger number of trials.

### Data analyses

#### Time course signal analysis

Time course signal analysis was conducted on the MEG data in order to isolate a time window that reflected the retention period in ASTM, devoid of contamination from other cognitive or sensory process (i.e., sensory or memory encoding or the motor anticipation of the test sequence). To this aim, we computed grand average waveforms for each condition (over trials and subjects). We used the signals from the Load 0 condition to set the boundaries of the time window used for later analyses of activity related to memory retention. The signal returned towards baseline (following stimulus presentation, see Fig. 1 in Supplementary material), and leveled off to a relatively constant level. As can be seen in Figs. 3D and E, the latency at which this first occurred varied across sensors, and we adopted an analysis window that started just beyond the longest leveling off latency in order to avoid contamination from the encoding phase. In order to avoid contamination from the motor anticipation of the test sequence, we set the final boundary at 100 ms before the presentation of the test sound sequence (Cheyne et al., 2006).
Source localization analysis

To perform source localization, the maximum entropy on the mean method (MEM, Amblard et al., 2004; Grova et al., 2006) was used on the time-varying MEG signals. This method uses a cortically-constrained distributed source-localization approach. The cortical surface (we used the white/gray matter boundary in a 3D surface) was segmented from each anatomical MRI scan using BrainVisa software (http://brainvisa.info/index_f.html, Mangin et al., 1995; Riviere et al., 2002). Approximately 4000 sources, oriented perpendicularly relative to the cortical surface, were distributed over the entire cortical surface, and these local sources were used in distributed source localization analyses, separately for each participant.

To perform a group analysis of the localization data, the following procedure was used:

1. We averaged the segmented and baseline-corrected signals over trials within each load condition, separately for interspersed and contiguous trials.
2. Next, we performed a MEM source localization on the cortical surface, for each sample point of the event-related magnetic field in the analysis window for memory maintenance (1100–1900 ms), for each of the 3 experimental conditions (Loads 0, 2, 4), separately for interspersed and contiguous trials, producing 960 localization images for each participant per condition. For each condition, we averaged the 960 localization images, and obtained an average cortically-constrained localization map for each condition, for each participant.
3. For each participant, the 3D average cortical-surface localization map of the MEG sources (see above) was converted into a volumetric 3D image. This surface-to-volume transformation was performed as follows:
   a) The intensity values for all voxels in a 3D volume (with the same dimensions as the original anatomical scan) were set to 0.
   b) The 3D surface was then positioned with the volumetric 3D image created in a.
   c) The voxels from the 3D volume that contained the sources from the 3D surface were attributed intensity values equal to those of the original sources.
4. The resulting images were smoothed using a Gaussian filter with an 8-mm FWHM (full width at half maximum).
5. Finally, all images were normalized to a common template in MNI-Talairach space using SPM2 (Ashburner & Friston, 1999).

Statistical analyses on source localization maps

We performed covariation analyses using a General Linear Model (AFNI software, Cox, 1996). We assumed that brain activity in the latter part of the retention interval would vary as a function of the number of items held in memory and therefore reflect brain activity related to the retention of basic acoustic attributes. To rule out the possibility that the presentation mode of the first sequence (interspersed vs. contiguous) still had an impact on brain activity in this latter part of the retention interval (e.g., reflecting a “re-organization” of the sequence in interspersed trials), a hierarchical regression approach was used. One model considered only the main effect of retained tones (K; Cowan, 2001) and the main effect of the mode of presentation of the tones (interspersed vs. contiguous), whereas the full model included both main effects and their interaction. That is, for each subject, each load (0, 2, or 4), and each presentation mode (interspersed or contiguous), localization images were regressed on the respective regressors of each hierarchical step for that subject. Analyses then included the following regressors:

1. the global mean,
2. the regression on K, the regression on presentation mode, and – only during the first step – the regression on the interaction between K and presentation mode,
3. and 11 regressors used for effect coding of the participant main effects (Cohen & Cohen, 1983).

For the main effect of ‘presentation mode’ (interspersed vs. contiguous, coded as −1 vs. 1 at Loads 2 and 4, and coded as 0 at Load 0) and the interaction of presentation mode with K (K = 0 at Load 0). We also performed an analysis that included only Loads 2 and 4, for which the coding of presentation mode was unambiguous.

The analyses revealed brain regions with activation patterns that have a significant linear relationship with the number of items held in memory. Clusters of voxels were deemed significant at p < 0.05 (corrected) according to the random field theory (Worsley et al., 2002). Each voxel in a significant cluster had a t-value ≥ 3.25 and a cluster volume ≥ 590 mm³. Brain activations in the cerebellum are not reported because the MEM method does not consider this region.

For a subject whose auditory memory span is greater than the memory load (i.e., number of tones in the sequence), the value of K (i.e., number of objects in auditory memory) is equal to the memory load. The more the memory load exceeds the subject’s memory span, the more K will differ from the load. In the event that K is equal (or very close) to the memory load then it follows that the above GLM analyses considering K or memory load, will reveal similar results. If the brain activity in one region is intimately correlated with K, the activity in this region must also be significantly correlated with variations in K (i.e. a deviation in K with respect to the memory load). In order to identify the brain areas that are significantly correlated with K, we must then test this correlation when K is highly variable and different than the load. In the present study, Load 4 represents the condition in which K will be most variable since this is the largest load and therefore most likely to exceed the subject’s memory span. Based on our data (see Fig. 2) it is clear that most of our subjects reached their memory span limit before or at load four (K at Load 4 was significantly lower than 4, p < .001, M = 1.94, SD = 1). In order to identify significant clusters of brain activity (among those previously found in the GLM analyses) that were correlated with the number of objects in auditory memory (and not only with the load), we searched for areas that showed a positive correlation between mean activity in the cluster (i.e., Load 4 activity minus Load 0 activity) and K values at Load 4, K(4).

Results

Behavioral results

Mean accuracy was 95% (SD = 8) for Load 0, 89.2% (SD = 11.4) for Load 2, and 83.2% (SD = 13) for Load 4. Mean accuracy decreased significantly with an increase in load, F(1, 12) = 29.98, MSE = 7.8, p < .001. In interspersed trials, the mean K value was 1.44 (SD = 0.46) and 1.78 (SD = 0.85) for Loads 2 and 4 respectively. In contiguous trials, the mean K value was 1.56 (SD = 0.43) and 2.10 (SD = 0.79) for Loads 2 and 4 respectively. A 2 × 2 ANOVA with the within-factors load (2, 4) and presentation mode (interspersed vs. contiguous) revealed significant main effects for presentation mode (p < .0001) and load (p < .0001) and a significant interaction between presentation mode and load (p < .0001).

Fig. 2. Behavioral results. Mean ASTM memory capacity, K, is plotted (continuous curve for contiguous trials, thick dashed line for interspersed trials) for the 13 participants as a function of the number of items to be encoded (load condition). Error bars correspond to standard errors. The diagonal thin dashed line shows the expected results for perfect memory.
Fig. 3. MEG time course results. A shows the grand average MEG recording for the memory Load 4 condition (mean signals calculated over all trials and subjects for all channels, green traces are for sensors on the left, blue for sensors on the right, light-blue for midline sensors). The map in B represents the topography of magnetic fields at the peak of the auditory N1 generated by the presentation of the second sound (seen from the top, anterior at the top, posterior at the bottom). The CTF convention (blue positive, indicates outgoing fields, red shows field lines entering the head) was used. The map in C represents the topography of magnetic fields during the retention period (same map orientation and CTF convention as B). D and E represent the mean MEG signal (average signal calculated over all trials and subjects) registered by sensors MLT32 and MRP57 for Loads 0, 2, and 4. The light-blue box in D and E corresponds to the time-window analysis for the retention period (1100 ms to 1900 ms). We filtered our data (40 Hz low pass filter) for display purposes only.
a significant main effect of load, $F(1, 12) = 11.48$, $MSE = 0.22$, $p < .01$, $\eta^2 = .49$; and a significant main effect of presentation mode, $F(1, 12) = 5.71$, $MSE = 0.11$, $p < .04$, $\eta^2 = .32$. The interaction between memory load and presentation mode was not significant, $F(1, 12) = 1.52$, $MSE = 0.09$, $p > .24$, $\eta^2 = .11$. Mean K values for each load are shown in Fig. 2, which also shows error bars corresponding to the standard error of the mean. Although the mean number of items held in memory increased as we presented more items to retain, there was considerable variability across individuals especially at Load 4.

**MEG results**

**Time course signal**

**Fig. 3A** shows the grand average MEG recording for all sensors at Load 4. The different phases of the experiment are shown from left to right: the presentation of the memory sequence (i.e., the first seven peaks corresponding to auditory N1 response to each sound), the retention interval, and finally the presentation of the test sequence (i.e., the final four auditory N1 peaks). **Fig. 3B** shows the magnetic field distribution of the auditory ERF (ERP MEG equivalent) N1 response for the second tone. The map in **Fig. 3C** represents the topology of magnetic fields during the retention period.

**Figs. 3D and E** show the grand average of the MEG signals for each load for two sensors showing clear load effects during the retention interval (MRT32 and MLP57). Again, the different phases during the trial sequence are clearly evident in the signals. Note how the signal recorded for the Load 0 condition reached a constant level following the encoding phase. After examining the time courses for all sensors at Load 0, we defined a time-window (see light-blue box in Figs. 3D and E) for the retention period starting at 1100 ms (corresponding to the point at which all sensors returned towards baseline, see Fig. 1 in Supplementary material) and ending at 1900 ms (corresponding to 100 ms before the test sequence). By choosing this time window, we minimized the possible contamination of activity related to the encoding of the first sequence or anticipation of the test sequence.

**Source localization maps**

The results of the source localization analyses focused on brain activity with amplitudes that varied with effective memory load (as measured by the estimated number of items held in memory during the task for each load for each participant). At the same time, we wanted to rule out the possibility that the presentation mode of the first sequence still had an impact on the brain activity during the latter part of the retention interval. In the first step, we looked for brain regions for which the interaction between presentation mode and memory capacity (K) was statistically significant. None were found (with or without taking Load 0 data into account). In the second step, we investigated if presentation mode explained any variance in brain activity, controlling for capacity, and again, there were no significant clusters (with or without taking Load 0 data into account). In the same regression, we investigated if memory capacity explained any variance in brain activity while controlling for presentation mode, and the analysis revealed 11 significant clusters. The results are shown in **Fig. 4** and summarized in Table 1. The 11 significant clusters at $p < .05$, corrected (cluster volume $\geq 590$ mm$^3$) were in the following brain areas: left and right superior parietal lobe (cluster 1 and cluster 2, BA 5/7), left inferior frontal gyrus/precentral gyrus (cluster 3, BA 6/9/44) and right middle/inferior frontal gyrus/precentral gyrus (cluster 4, BA 6/9/44), left middle temporal gyrus (cluster 5, BA 39) and right middle/inferior temporal gyrus (cluster 6, BA 37/39), left and right superior/middle temporal gyrus (cluster 7 and cluster 8, BA 21/22), left superior/middle frontal gyrus (cluster 9, BA 9/10/46, and cluster 10, BA 10), and right parahippocampal gyrus (cluster 11, BA 35, see also Fig. 1 in the Supplementary material).

Among the 11 clusters revealed by the GLM (see Table 1), we investigated the brain regions that showed a significant positive correlation between mean activity in the cluster (i.e., Load 4 activity minus Load 0 activity) and K values at Load 4, K(4). Such clusters were found in the right middle/inferior frontal gyrus/precentral gyrus (cluster 4, BA 6/9/44, $r(24) = .55$, $p < .01$), in the right middle/inferior temporal gyrus (cluster 6, BA 37/39, $r(24) = .49$, $p < .02$), in the right
superior/middle temporal gyrus (cluster 8, BA 21/22, r\(24\) = .46, \(p < .02\)), and in the right superior/middle frontal gyrus (cluster 10, BA 10, r\(24\) = .64, \(p < .001\)). Cluster 2 (right superior parietal lobe, BA 5/7, r\(24\) = .35, \(p < .08\)), cluster 3 (left inferior frontal gyrus/precentral gyrus, BA 6/9/44, r\(24\) = .45, \(p < .06\)), cluster 7 (left superior/middle temporal gyrus, BA 21/22, r\(24\) = .38, \(p < .06\)), cluster 9 (left superior/middle frontal gyrus, BA 9/10/46, r\(24\) = .33, \(p < .1\)), and cluster 11 (right parahippocampal gyrus, BA 35, r\(24\) = .35, \(p < .1\)) showed a non-significant trend (see Table 2 and Fig. 5).

**Discussion**

In the present study, we sought to improve our understanding of the neural basis of ASTM in the human brain. To do so, we used MEG to examine brain activity related to the maintenance of pitch information in ASTM. We focused on brain activity that increased with the number of retained items during the retention interval of a memory task, of the same regions as found in the present study, namely in the inferior temporal and spatial resolution of the MEG technique (i.e., relative to EEG), permitted a more precise localization of brain activity related to ASTM. The present study is the first to use MEG to provide both fine temporal and spatial localization information on the neural mechanisms of non-verbal ASTM for pitch in relation to individual-differences in memory capacity.

**Brain activity in frontal and temporal cortices predicts individual differences in ASTM capacity**

Our results are generally consistent with findings from previous brain imaging studies on non-verbal ASTM. More specifically, using PET, Zatorre et al. (1994) found brain activation related to non-verbal ASTM for pitch in similar brain regions as found in the present study, namely the superior/middle temporal gyrus (BA 21/22) and the middle frontal gyrus (BA 9/10/46). Griffiths et al. (1999) also reported brain activity in response to ASTM in the middle frontal gyrus. Using fMRI, Gaab et al. (2003) found brain activity related to non-verbal ASTM for pitch in these same areas frontal and temporal regions, and also in the inferior frontal gyrus (BA 9/10/45), very close to an area found in the present study. Most recently, using fMRI, Koelsch et al. (2009) found brain activity in response to the rehearsal of non-verbal ASTM for pitch in some of the same regions as found in the present study, namely in the inferior frontal gyrus (BA 9/44/45) and middle frontal gyrus (BA 9/10/46). Taken together, results from the present study and previous brain

**Table 1**

Covariation analysis of brain activity and K. The table shows brain areas where brain activity increased with an increase in the number of tones maintained in ASTM in a group analysis. This analysis was performed to reveal brain regions with activation patterns that have a significant linear relationship with the number of items effectively held in memory in the source localization maps, significant at \(p < .05\) corrected (\(df = 51\), see Methods and Results sections for further details).

<table>
<thead>
<tr>
<th>Cluster number</th>
<th>Brain area</th>
<th>Brodmann area (BA)</th>
<th>Hemisphere</th>
<th>Sterotaxic coordinates (MNI-Tal space) for center of cluster</th>
<th>Volume of cluster (mm(^3))</th>
<th>(t)-Value for peak voxel in cluster ((df = 51))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Superior parietal lobe</td>
<td>BA 5/7</td>
<td>Left</td>
<td>(x = -29), (y = -50), (z = 66)</td>
<td>7560</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>Superior parietal lobe</td>
<td>BA 5/7</td>
<td>Right</td>
<td>(x = -18), (y = -54), (z = 18)</td>
<td>7896</td>
<td>9.15</td>
</tr>
<tr>
<td>3</td>
<td>Inferior frontal gyrus, precentral gyrus</td>
<td>BA 6/9/44</td>
<td>Left</td>
<td>(x = -63), (y = -8), (z = 20)</td>
<td>4808</td>
<td>10.5</td>
</tr>
<tr>
<td>4</td>
<td>Middle/inferior frontal gyrus, precentral gyrus</td>
<td>BA 6/9/44</td>
<td>Right</td>
<td>(x = 49), (y = 1), (z = 21)</td>
<td>9280</td>
<td>8.7</td>
</tr>
<tr>
<td>5</td>
<td>Middle temporal gyrus</td>
<td>BA 39</td>
<td>Left</td>
<td>(x = -49), (y = -77), (z = 10)</td>
<td>5528</td>
<td>8.7</td>
</tr>
<tr>
<td>6</td>
<td>Middle/inferior temporal gyrus</td>
<td>BA 37/39</td>
<td>Right</td>
<td>(x = -49), (y = -69), (z = 2)</td>
<td>1240</td>
<td>8.1</td>
</tr>
<tr>
<td>7</td>
<td>Superior/middle temporal gyrus</td>
<td>BA 21/22</td>
<td>Left</td>
<td>(x = -63), (y = -24), (z = -15)</td>
<td>4352</td>
<td>9.1</td>
</tr>
<tr>
<td>8</td>
<td>Superior/middle temporal gyrus</td>
<td>BA 21/22</td>
<td>Right</td>
<td>(x = 59), (y = -21), (z = -12)</td>
<td>4240</td>
<td>9.6</td>
</tr>
<tr>
<td>9</td>
<td>Superior/middle frontal gyrus</td>
<td>BA 9/10/46</td>
<td>Left</td>
<td>(x = -34), (y = 44), (z = 24)</td>
<td>7072</td>
<td>9.5</td>
</tr>
<tr>
<td>10</td>
<td>Superior/middle frontal gyrus</td>
<td>BA 10</td>
<td>Right</td>
<td>(x = 15), (y = 61), (z = 4)</td>
<td>2536</td>
<td>8.5</td>
</tr>
<tr>
<td>11</td>
<td>Parahippocampal gyrus</td>
<td>BA 35</td>
<td>Right</td>
<td>(x = 22), (y = -24), (z = -20)</td>
<td>832</td>
<td>8.2</td>
</tr>
</tbody>
</table>

**Table 2**

Correlation between K(4) and the difference of brain activity at Load 4 and Load 0, in selected clusters. The table shows the correlation between K at Load 4 and the difference of Load 4 activity and Load 0 activity in the 11 clusters that have been revealed by the covariation analysis. The analysis revealed a significant positive correlation at 4 right-hemispheric clusters and a marginally significant positive correlation at 5 additional clusters in both hemispheres (\(df = 24\) for each analysis).

<table>
<thead>
<tr>
<th>Cluster number</th>
<th>Brain area</th>
<th>Brodmann area (BA)</th>
<th>Hemisphere</th>
<th>(p) value for the regression: activity increase (Load 4 minus Load 0) vs K(4) ((df = 24))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Superior parietal lobe</td>
<td>BA 5/7</td>
<td>Left</td>
<td>0.5562 (p &lt; .05)</td>
</tr>
<tr>
<td>2</td>
<td>Superior parietal lobe</td>
<td>BA 5/7</td>
<td>Right</td>
<td>0.0764 (p &lt; .01)</td>
</tr>
<tr>
<td>3</td>
<td>Inferior frontal gyrus, precentral gyrus</td>
<td>BA 6/9/44</td>
<td>Left</td>
<td>0.0543 (p &lt; .05)</td>
</tr>
<tr>
<td>4</td>
<td>Middle/inferior frontal gyrus, precentral gyrus</td>
<td>BA 6/9/44</td>
<td>Right</td>
<td>0.0038 (p &lt; .01)</td>
</tr>
<tr>
<td>5</td>
<td>Middle temporal gyrus</td>
<td>BA 39</td>
<td>Left</td>
<td>0.2319 (p &lt; .1)</td>
</tr>
<tr>
<td>6</td>
<td>Middle/inferior temporal gyrus</td>
<td>BA 37/39</td>
<td>Right</td>
<td>0.0113 (p &lt; .05)</td>
</tr>
<tr>
<td>7</td>
<td>Superior/middle temporal gyrus</td>
<td>BA 21/22</td>
<td>Left</td>
<td>0.0580 (p &lt; .05)</td>
</tr>
<tr>
<td>8</td>
<td>Superior/middle temporal gyrus</td>
<td>BA 21/22</td>
<td>Right</td>
<td>0.0184 (p &lt; .05)</td>
</tr>
<tr>
<td>9</td>
<td>Superior/middle frontal gyrus</td>
<td>BA 9/10/46</td>
<td>Left</td>
<td>0.0943 (p &lt; .05)</td>
</tr>
<tr>
<td>10</td>
<td>Superior/middle frontal gyrus</td>
<td>BA 10</td>
<td>Right</td>
<td>0.0004 (p &lt; .05)</td>
</tr>
<tr>
<td>11</td>
<td>Parahippocampal gyrus</td>
<td>BA 35</td>
<td>Right</td>
<td>0.0901 (p &lt; .05)</td>
</tr>
</tbody>
</table>
imaging studies (Gaab et al., 2003; Koelsch et al., 2009; Zatorre et al., 1994) converge on frontal and temporal cortices as playing a critical role in non-verbal ASTM for pitch.

While our results mostly converge with the above previous brain imaging findings on non-verbal ASTM, there are some differences across these studies. First, we found additional activation in frontal cortex including the precentral gyrus (cluster 3 and cluster 4, BA 6/9/44) that was not found in the above previous studies on non-verbal ASTM for pitch. These discrepancies across our study and previous work on non-verbal ASTM for pitch may be due to differences in experimental design. In the present study, we were able to isolate brain activity occurring uniquely during the retention period (after encoding, and prior to retrieval, comparison, and response-selection and execution operations required by the task), and to isolate brain regions that were correlated to variations in ASTM load. Thus, our results establish a tighter relationship between ASTM function and observed brain activity. We provide evidence of specific brain regions, the right middle/inferior frontal gyrus/precentral gyrus (cluster 4, BA 6/9/44), the right middle/inferior frontotemporal cortex, and the right parahippocampal gyrus.

Fig. 5. Relationship between the increase of the mean activity and the number of objects in auditory memory for each cluster. The figure shows, for clusters 4, 6, 8, and 10, a significant positive correlation between the increase of the mean activity (activity at Load 4 minus activity at Load 0) and K at Load 4, K(4) (p < .05, marked with two asterisks). The correlation was marginally significant for clusters 2, 3, 7, 9, and 11 (p < .1, marked with one asterisk). Note that the scaling on the ordinate of cluster 11 is different from the scaling of the other clusters. For further details see Table 2.
temporal gyrus (cluster 6, BA 37/39), the right superior/middle temporal gyrus (cluster 8, BA 21/22), in the right superior/middle frontal gyrus (cluster 10, BA 10), and, at a marginally significant level, the right superior parietal lobe (cluster 2, BA 5/7), the left inferior frontal gyrus/ precentral gyrus (cluster 3, BA 6/9/44), the left superior/middle temporal gyrus (cluster 7, BA 21/22), the left superior/middle frontal gyrus (cluster 9, BA 9/10/46), and the right parahippocampal gyrus (cluster 11, BA 35) that predicted individual differences in ASTM capacity, wherein brain activity was greater in magnitude in individuals with higher ASTM capacity. Thus, while the frontal, parietal, and temporal cortices appear to be involved in non-verbal ASTM in general, these regions also predict individual differences in ASTM capacity.

Finally, our findings are consistent with ERP studies of non-verbal ASTM conducted in our laboratory (Guimond et al., 2011; Lefebvre et al., 2013; Nolden et al., 2013). We note that Guimond et al. (2011) presented either a single pitch or two pitches, simultaneously, as the main manipulation of load. This enabled the study of a load-related response using simultaneous presentations of the stimuli, unlike the present approach that used a sequential presentation. Despite these differences, Guimond et al. (2011) found a frontal–central brain response that varied with increasing ASTM load. Thus, the present MEG study went beyond our earlier ERP work by providing precise spatial localization of brain activity related to ASTM, and by showing that brain activity correlates with individual differences in ASTM capacity.

**Implications for models of STM**

Previous models of STM (Baddeley, 1986, 2001; Baddeley & Hitch, 1974) proposed that dedicated memory systems maintain information from different sensory inputs in brain areas that are distinct from those that mediate perception, and that a ‘central executive’ system coordinates the flow of information through the system. According to Baddeley, there are two main components involved in ASTM: 1) a “phonological loop” that involves an auditory memory store and Articulatory rehearsal processes, as well as 2) a “central executive” that manages and manipulates the information. Our GLM analyses revealed 11 clusters that were both load and K dependent. In four of these clusters, we observed a significant correlation between the change in activity with higher memory load and the average number of items retained during condition Load 4, particularly in right frontal and temporal regions (see Fig. 5). If we consider that these frontal peaks correspond to the “central executive,” it follows that the latter would thus be load and K dependent. In contrast to the present findings in ASTM, similar studies of visual STM (including several studies from our laboratory that used a similar approach, e.g. Grimault et al., 2009b; Robitaille et al., 2009; Todd & Marois, 2004), did not reveal any load dependent frontal regions thought to correspond to a central executive. This suggests that central executive functioning may differ across the auditory and visual STM domains. Baddeley’s theory does not speak of different “central executives” for each modality (e.g., auditory or vision). Thus, if we wish to associate our ASTM load-dependent frontal activation patterns with a central executive, given that the central executive in VSTM does not appear to be similarly load-dependent, this would entail a major revision to the Baddeley model. If rather, our results are linked to a “phonological loop,” an alternative hypothesis might be that the four frontal clusters, cluster 3 (left inferior frontal gyrus/precentral gyrus, BA 6/9/44), cluster 4 (right middle/inferior frontal gyrus/precentral gyrus, BA 6/9/44), cluster 9 (left superior/middle frontal gyrus, BA 9/10/46), and cluster 10 (right superior/middle frontal gyrus, BA 10) are part of an “auditory memory store” (since they are more closely related to K) and that the other clusters may be associated with a “covert rehearsal process” (i.e., imagining the sequence in the mind’s ear or covert singing). Note, however, that articulatory suppression did not strongly interfere with tone retention in a similar paradigm (see Guimond et al., 2011, Exp. 3).

In contrast to Baddeley’s model, other models (Cowan, 2008; D’Esposito, 2007; D’Esposito et al., 2000; Goldman-Rakic, 1987; Petrides, 1991, 2005; Postle, 2006; Postle et al., 1999; Ruchkin et al., 2003) propose that memory maintenance occurs in brain areas that support sensation and perception. These models predict that ASTM for pitch should implicate pitch-processing centers in the brain. In support of this latter approach, here we found brain activity in previously found pitch-relevant brain areas including the left and right superior/middle temporal gyri (clusters 7 and 8, BA 21/22) and left and right superior/middle frontal gyri (clusters 9 and 10, BA 9/10/46, e.g. Zatorre et al., 1994).

While most clusters found in this study were found in symmetrically-homologous regions (with the exception of cluster 11, the right parahippocampal gyrus, BA 35), there was a slight advantage of the right hemisphere in that the correlation between individual memory capacity and changes in activity was more pronounced in the right than in the left hemisphere. If laterality is not considered, areas that show increased activity when the number of retained items increases are relatively consistent with those of Rauschecker and Scott (2009) (see Fig. 5). In particular, our clusters 7 and 8 (left and right superior/middle temporal gyr, BA 21/22) are very close to an area of the auditory cortex suggested to be the separation point of the auditory “what” and “where” streams. Moreover, our clusters 3 and 4 (left and right middle frontal gyrus/precentral gyrus, BA 6/9/44) correspond to the end of the “what” stream, while also partially overlapping with the end of the “where” stream. In our task, there was no explicit “where” component, so that overlapping activity with the “where” stream seems surprising at first. Alternatively, activations found in the middle frontal gyrus could be part of a “when” stream which would be in line with the temporal processing required by our task. Indeed, some studies (Crozier et al., 1999; Knutson et al., 2004) have reported that the middle frontal gyrus (BA 6, BA 8) is implicated bilaterally in the processing of sequences, with an advantage of the left hemisphere over the right hemisphere. Rauschecker and Scott (2009) model was designed to summarize results from speech perception and production and considers these to be left-lateralized. However, some authors have questioned the laterality scheme of the Rauschecker and Scott (2009) model. For example, it has been shown that the brain works more bilaterally to perceive speech sounds and that the left part is more closely related to semantically-based recognition (Hickok et al., 2008; Rogalsky et al., 2008). In our protocol, we used pure tones with frequencies selected to be ‘non-musical’ to minimize semantic activation or activations based on long-term memory; furthermore, no sound production was required. Following the work of Hickok et al. (2008) and Rogalsky et al. (2008), the “non-semantic” perception of our stimuli could be done in both hemispheres. Moreover, in our protocol, participants needed to process pitch, which, according to Zatorre et al. (1994), would be performed with a slight advantage of the right frontal cortex. The relatively stronger implication of the right hemisphere in our task may also be explained by the recent idea that the right auditory cortex is specialized to process fine spectral information (for a review, see Zatorre et al., 2002). As our task requires pitch processing, it may rely more strongly on the right than on the left auditory cortex and in consequence more strongly engage related right hemispheric areas without turning down left hemispheric activity. In summary, our findings of activity in the left and right superior/middle temporal gyr (clusters 7 and 8, BA 21/22) might reflect perceptual-based processing, while the closer relation of right hemispheric activations to memory capacity might reflect a relative advantage of the right hemisphere for spectral processing. Then, based on a memory model that considers that memory retention requires the reactivation of brain areas involved in the initial perception of the stimulus (Cowan, 2008; D’Esposito, 2007; D’Esposito et al., 2000; Goldman-Rakic, 1987; Kosslyn, 1980; Petrides, 1991; Postle, 2006; Postle et al., 1999; Ruchkin et al., 2003), the activation of the frontal areas would be sustained during retention. In summary, it is possible that the clusters we identified could be part of
Oscillatory activity

In the present study we averaged the signal over long periods of time, effectively examining a low-passed representation of the signal. Some authors have studied oscillatory activities during auditory memory tasks (e.g., Kaiser et al., 2003; Luo et al., 2005; Lutzenberger et al., 2002, see also Albouy et al., 2013). Lutzenberger et al. (2002) investigated oscillatory magnetoencephalographic activity during a memory task for auditory spatial information. Their auditory spatial working memory task led to an increase in parietal activity in the gamma band (at 59 Hz) and an increased parieto-frontal coupling. They suggested that these latter findings might indicate an activation of the “frontal executive” and enhanced cross-talk between frontal and posterior “slave systems” (Baddeley, 1986, 2001; Baddeley & Hitch, 1974) during the “delay phase” (phase between stimuli) of the memory task. However, Lutzenberger et al. (2002) recognized that their data did not allow them to determine whether the representation of the stimuli to be maintained in memory took place in frontal or parietal areas because gamma activity did not show a time course that was strictly specific to the delay phase.

Kaiser et al. (2003) used the same type of protocol to investigate an auditory pattern memory task. The authors found an increase of the gamma band coherence between left fronto-temporal and prefrontal sensors throughout the delay period of the memory task. Our protocol could induce this type of activity (gamma band, parieto-frontal/ fronto-temporal-prefrontal) but there are major differences between our protocols and goals that would likely create differences in our results. The protocol used by these authors (Kaiser et al., 2003; Luo et al., 2005; Lutzenberger et al., 2002) did not include any parametric manipulation of memory load and therefore they could not investigate a correlation between the increase in activity and the number of objects held in memory. Luo et al. (2005) used a delayed-match-to-sample paradigm to study the perception and categorization of speech and non-speech sounds. The authors found alpha band activity in the auditory cortex, for both auditory discrimination and category discrimination conditions, during the delay period for all stimulus types. Luo et al. (2005) observed a difference between the auditory discrimination and category discrimination conditions for the non-speech stimuli in auditory areas, and for both speech and non-speech stimuli in frontal areas. Our work based on event-related averaging converges with many of these latter results (e.g., the implication of temporal and frontal regions in memory processes) and paves the way for future studies based on time-frequency methods that would include parametric variations in memory load.

Conclusions

Our results provide a systematic examination of ASTM retention activity while reducing possible confounding contributions from other sensory and cognitive processes (in particular, phonological, semantic, or linguistic mechanisms) to a minimum. Load-related MEG responses revealed a set of brain regions (particularly in lateral frontal and temporal areas) that were sensitive to memory load and to individual differences in number of items represented in ASTM. These results suggest a strong functional link between these regions and ASTM for basic, non-verbal, attributes of auditory stimuli, and contribute to our general understanding of how brain function is related to behavior in auditory memory for pitch.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neuroimage.2014.03.020.

Acknowledgments

We thank Nathalie Bouloute and Ping Hei Lam for technical assistance related to MEG data acquisition, Carrolyn Hurst, André Cyr, and Mathieu Desrosiers (Unité de Neuroimagerie Fonctionnelle, Université de Montréal), and Marc Bouffard (Montreal Neurological Institute, McGill University) for technical assistance related to the MRI data acquisition and processing. This work was made possible by support from the Canada Research Chairs (CRC) program and research grants from the Canadian Institutes of Health Research (CIHR), the Fonds de Recherche du Québec Nature et Technologies, and the Canada Foundation for Innovation awarded to Pierre Jolicoeur, and by infrastructure support from the Fonds de Recherche du Québec (FRSQ) and Université de Montréal.

References


