MANUSCRIPT ACCEPTED FOR PUBLICATION BY HUMAN BRAIN MAPPING JAN 2008

Functional Architecture of Verbal and Tonal Working Memory: An fMRI Study

Stefan Koelsch[1,2], Katrin Schulze[1], Daniela Sammler[1], Thomas Fritz[1], Karsten Müller[1]
& Oliver Gruber[3]

[1] Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

[2] University of Sussex, Brighton, UK

[3] Georg August University, Goettingen, Germany

ABSTRACT

This study investigates the functional architecture of working memory (WM) for verbal and tonal information during rehearsal and articulatory suppression. Participants were presented with strings of four sung syllables with the task to remember either the pitches (tonal information) or the syllables (verbal information). Rehearsal of verbal, as well as of tonal information activated a network comprising ventrolateral premotor cortex (encroaching Broca's area), dorsal premotor cortex, the planum temporale, inferior parietal lobe, the anterior insula, subcortical structures (basal ganglia and thalamus) as well as the cerebellum. The topography of activations was virtually identical for the rehearsal of syllables and pitches, showing a remarkable overlap of the WM components for the rehearsal of verbal and tonal information. When the WM task was performed under articulatory suppression, activations in those areas decreased, while additional activations arose in anterior prefrontal areas. These prefrontal areas might contain additional storage components of verbal and tonal WM that are activated when auditory information cannot be rehearsed. As in the rehearsal conditions, the topography of activations under articulatory suppression was nearly identical for the verbal as compared to the tonal task. Results indicate that both the rehearsal of verbal and tonal information, as well as storage of verbal and tonal information relies on strongly overlapping neuronal networks. These networks appear to partly consist of sensorimotor-related circuits which provide resources for the representation and maintenance of information, and which are remarkably similar for the production of speech and song.

Keywords: Working Memory, pitch, verbal, music

INTRODUCTION

Working memory (WM) refers to a brain system of linked and interacting information-processing components for temporal storage and simultaneous manipulation of information (Baddeley, 1992; Baddeley, 2003). This brain system is critical for higher cognitive functions such as language, music, planning, problem solving and reasoning. One of the most influential WM models was developed by Baddeley & Hitch more than three decades ago (Baddeley & Hitch, 1974; Baddeley, 1992). According to this model, WM consists of an attentional control system (the "*central executive*") that operates in conjunction with two "*slave systems*" that serve to maintain representations of information of different modalities: the visuospatial sketchpad and the phonological loop. The visuospatial sketchpad is concerned with the processing and storage of visual and spatial information. The phonological loop represents verbal short-term memory, and is thought to consist on the one hand of a phonological store that holds auditory information for a few seconds, and on the other hand of a phonological rehearsal mechanism that is analogous to subvocal speech (Baddeley, 2003).

So far, the functional neuroarchitecture of the phonological loop has mainly been investigated with respect to language. Both neuropsychological and functional imaging studies indicate that Broca's area and premotor areas (pre-SMA, SMA, vlPMC and dPMC) play a crucial role during the phonological rehearsal process (Awh et al., 1996; Fiez et al., 1996; Gruber & von Cramon, 2003; Paulesu et al., 1993; Ravizza et al., 2004). In addition, both the insular cortex (Bamiou et al., 2003; Paulesu et al., 1993; Chein et al., 2002) and the cerebellum (Chen & Desmond, 2005; Gruber, 2001; Kirschen et al., 2005; Ravizza et al., 2004) have been reported to be involved in phonological rehearsal. The phonological store has been suggested to be located in parietal areas, particularly the inferior parietal lobe (Awh et al., 1996; Chen & Desmond, 2005; Crottaz-Herbette et al., 2004; Gruber, 2001; Gruber & von Cramon, 2003; Henson et al., 2000; Jonides et al., 1998; Kirschen et al., 2005; Paulesu et al., 1993), but also in the superior parietal lobe (Awh et al., 1996; Chen & Desmond, 2005; Crottaz-Herbette et al. 2004; Henson et al., 2000; Ravizza et al., 2004). However, the localization of the phonological store in the parietal lobe is partly controversial (Fiez et al., 1996; Hickok, Buchsbaum, Humphries, & Muftuler, 2003), because, e.g., neural activity in this area might also reflect increased engagement of attentional resources (for an overview see: Cabeza and Nyberg, 2000; Corbetta and Shulmann, 2002, see also Jones et al., 2004). Moreover, a series of recent functional neuroimaging and experimental neuropsychological studies has provided evidence that phonological storage is not a purely parietal brain function, but relies on a broader network of inferior parietal and anterior prefrontal brain regions supporting the non-articulatory maintenance of phonological information (Gruber, 2001; Gruber & von Cramon, 2001, 2003; Gruber & Goschke, 2004; Gruber et al., 2005).

Obviously, phonological information is not the only important auditory information in everyday life. Other relevant information includes speech prosody, and music. So far, a number of behavioural studies investigated whether the phonological loop also serves the processing of such non-phonological information, or whether different sub-systems (like a "*tonal loop*", see Pechmann & Moor, 1992) exist in addition to the phonological loop. However, these studies do not yet provide a consistent picture. Deutsch (1970) reported that intervening tones interfered more strongly than phonemes with a pitch memory task, and this finding was taken as evidence for a specialized tonal WM system. Salame & Baddeley (1989) showed that vocal music interfered more strongly with phonological short-term memory than instrumental music, supporting the assumption of two independent WM systems for verbal and tonal stimuli (Salame & Baddeley, 1989). On the other hand, results by Semal, Demany, Ueda, and Halle (1996) suggest that the pitch of speech sounds is not stored differently from the pitch of non-speech sounds in WM. In addition, Iwanaga and Itoh (2002) reported that instrumental as well as vocal music interfered with a verbal WM task, and Chan et al. (1998) observed that musical training increases the performance during a verbal WM task, suggesting rather overlapping neural resources for verbal and tonal WM. Considering these contradictory results, it remains unclear whether cognitive (and neural) resources of tonal and verbal WM overlap. Thus, knowledge about the neural organization of the phonological loop cannot simply be generalized to non-phonological auditory WM.

Possible differences or similarities between the neuronal networks underlying WM for tonal and verbal stimuli have so far only sparsely been addressed. Using fMRI, Gaab, Gaser et al. (2003) showed involvement of the supramarginal gyrus (SMG) extending into the intraparietal sulcus (IPS), planum temporale, ventrolateral premotor regions encroaching Broca's area, dorsolateral premotor regions, and dorsolateral cerebellar regions during a pitch memory task. This network is surprisingly reminiscent of the network implicated in the phonological loop (see above). A similar network (including the inferior frontal and insular cortex, the planum temporale, and the SMG) had previously been shown with PET for the active retention of pitch (Zatorre et al., 1994). Hickok, Buchsbaum, Humphries and Muftuler (2003) showed with fMRI that (subvocal) articulatory rehearsal of verbal as well as of musical information activated ventrolateral premotor regions encroaching Broca's area, dorsolateral premotor regions, the planum temporale (referred to by the authors as area Spt), and (with lowered statistical threshold) the SMG/IPS. The regions activated in that study were, thus, very similar compared to those observed by Gaab et al. (2003), and they served the processing of both verbal and musical rehearsal.

In the present study, we investigated similarities and differences between the neural components underlying WM for verbal (syllables) and tonal (pitch) material both during rehearsal and under articulatory suppression. The stimuli for the tonal and verbal WM tasks were identical, consisting of sequences of four sung syllables. To investigate the articulatory rehearsal component, participants were required to rehearse subvocally either the pitches or the syllables after the presentation of a stimulus sequence, and to respond subsequently to a probe sequence with a button press. In addition, suppression conditions were employed to assess the neural correlates of the non-articulatory storage component. Articulatory suppression is known to prevent articulatory rehearsal, and participants therefore have to rely more strongly on the information represented in the phonological store to perform the task (Gruber, 2001). During the suppression condition in the present experiment, participants were asked to remember either the pitches or the syllables of a presented sequence, while singing a well known children's song after the presentation of the sequence. Only this combined articulatory and musical (tonal) suppression task was able to prevent the subjects from using a tonal rehearsal strategy, i.e. to subvocally repeat the pitches while performing (purely) articulatory suppression. After the suppression, participants were asked to respond subsequently to a probe sequence with a button press (as in the rehearsal conditions).

Based on the literature reported above, we hypothesized that articulatory rehearsal would activate frontal speech areas (ventral premotor cortex and Broca's area), parietal regions (SMG/IPS), and the planum temporale. Region of interest analyses were planned to investigate possible differences in the topography of the activated networks for tonal and verbal rehearsal, as well as hemispheric differences and differences in strength of activation. In the suppression conditions, additional activations were expected in anterior prefrontal (intermediate frontal sulcus) and inferior parietal areas.

METHODS

Participants. Twelve right-handed non-musicians (25 to 30 years, *M* = 26.7 years, 7 females) with normal hearing took part in the experiment. None of the participants had any special musical training exceeding general school education. All subjects were students of the University of Leipzig (except one, who was a chef in a restaurant). Participants were right-handed (mean lateralization quotient was 95.8% according to the Edinburgh Handedness Inventory; Oldfield, 1971), and had reading span scores ranging from 2.6 to 6 (*M* = 3.7, *SD* = 1.1; scores were assessed with a German version of the reading span test from Daneman & Carpenter, 1980).

Stimuli. Stimuli were sung syllables, thus containing both verbal (syllable) and tonal (pitch) information (no spoken syllables, and no pure tones or instrumental tones were presented). There were eight syllables (taken from the German alphabet) which were acoustically well distinguishable (b [be;], f [Ef], j [jOt], k [ka;], o [o;], v [fAu], x [Iks] and z [tsEt]). Each of these eight syllables was sung by a female singer on eight different pitches (these eight pitches corresponded to the pitches of a major scale), resulting in a total of 64 sung stimuli (8 syllables x 8 pitches = 64). The pitches of the stimuli were electronically adjusted using *Cool Edit Pro* (Syntrillium Corp., Phoenix, Arizona / USA) within a range from 200 to 400 Hz (corresponding to one octave) with interval ratios exactly corresponding to tempered intonation. Length of stimuli was adjusted to 400±2 ms by shortenting vowels only (thus without reducing intelligibility of the syllables). To construct control conditions, each stimulus was also recorded backwards (see also below). Subsequently, stimuli were grouped to 216 sequences, each comprising of four stimuli (Figure 1). Syllables did not form meaningful words (such as "fox") and pitches of consecutive stimuli were at least five, and not more than nine, semitones apart from each other. Silence periods of 150 ms were inserted between stimuli, and a 100 ms pause was added after the last stimulus.

Procedure. There were six experimental conditions (see also Figure 1): (1) memorize pitches while rehearsing the pitches, (2) memorize pitches during articulatory suppression (singing a children's song, see also below), (3) memorize syllables while rehearsing the syllables, (4) memorize syllables during articulatory suppression (singing the children's song). In addition to these memory conditions, there were two control conditions: (5) memorize nothing (without rehearsal or singing), and (6) memorize nothing and sing the children's song.

Each experimental trial started with a visual cue consisting of two simultaneously presented capital letters (2350 ms). The first letter indicated what to memorize (see green letters in the left of Figure 1): either only syllables (i.e., the verbal information, "V"), or only pitches (i.e., the tonal information, "T"), or nothing ("N"). The second letter indicated the articulatory action to be performed after the presentation of a stimulus sequence (see red letters in the left of Figure 1): either rehearsal (of pitches or syllables, "R"), or singing the children's song ("S", this task also served the articulatory suppression, see below), or neither sing nor rehearse ("0"). For the singing condition, participants were instructed to subvocally sing a well-known German children's song ("Hänschen klein"). Importantly, during trials in which pitches or syllables had to be memorized, the singing condition represented a combined articulatory and musical (tonal) suppression because it prevented both tonal and verbal rehearsal (note that the material used for articulatory suppression was, thus, identical for the verbal and the tonal condition). For the rehearsal conditions, participants were instructed to subvocally rehearse the syllables (without melody) in the verbal condition, and in the tonal condition to subvocally rehearse the pitches (without articulating the syllables of the stimulus sequence) using the syllable [hm]. That is, the rehearsal task was designed such that participants only rehearsed pitches (without syllables), or only rehearsed syllables (without pitches). We used covert production to avoid auditory feedback of the subject's own voice (and corresponding activations of the auditory cortex), and to avoid motion artefacts in the fMRI signal that are likely to occur during overt oral production (see also Callan et al., 2006). Moreover, the potential interaction between the degree of susceptibility artefact related to changes in the oral cavity during scanning of vocal articulation on the one side, and the type of production task (verbal or tonal rehearsal) on the other, is likely to produce false results (see also Callan et al., 2006). However, participants were thoroughly trained with the tasks in a separate training session with both overt and covert production (see also below).

The initial cue of each trial was followed by a four-stimulus sequence that had a duration of 2150 ms, and by a silence period (4000 ms) for subvocal rehearsal or singing/suppression. After this period, participants were presented with a probe sequence that consisted, as the initial sequence, of four sung syllables (2150 ms). Then, subjects were asked to indicate via a button press whether the pitches of that sequence (in the tonal condition) or the syllables of that sequence (in the verbal condition) were the same as those of the initial sequence. As mentioned above, participants did not have to memorize syllables or tones in the two control conditions (N 0, N S).

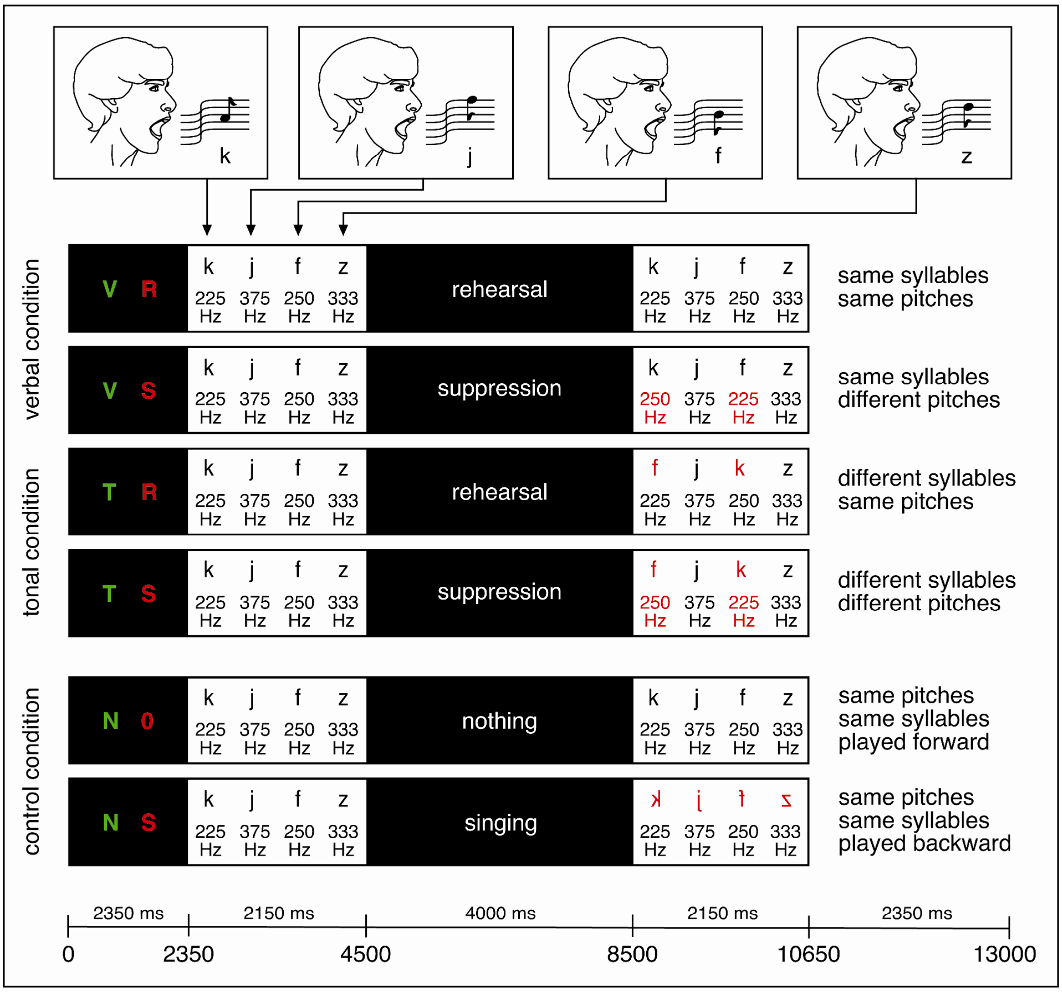

Figure 1: Experimental design. The six example trials illustrate the six different experimental conditions, each trial had a duration of 13 s and began with a visual cue (V R = verbal (syllable) rehearsal, V S = verbal (syllable) suppression, T R = tonal (pitch) rehearsal, T S = tonal (pitch) suppression, N 0 = not memorize anything without rehearsing or singing a song, N S = not memorize anything and sing the song). The cue was followed by the presentation of the stimulus sequence. In the subsequent rehearsal conditions, subjects covertly rehearsed either the syllables (V R) or the pitches (T R). During the suppression conditions (V S, T S), subjects covertly sang a children’s song while trying to maintain either the verbal (V S), or the tonal information (T S) in their memory. Then, a probe sequence was presented, followed by a silence period of 2.35 s during which participants had to indicate whether the probe sequence was identical to the initial sequence (verbal and tonal conditions). In the control conditions (N 0, N S), in which participants did not have to memorize the initial sequence, participants had to indicate whether each of the syllables was played forward (forward speech) or backward (backward speech).

In these conditions, stimuli of the probe sequence were played with the same pitches, but each of the syllables was either played forward or backward, and subjects had to indicate via a button press whether they heard forward speech or backward speech (see outermost right of the two bottom panels of Figure 1). This task was easy (correct responses were > 98 %, see Results), and although participants had nothing to remember, they still anticipated a control sequence, made a binary (yes/no) decision, and performed a motor response.

In the other conditions, four different types of probe sequences were used (see two outer most right panels in Figure 1): (a) verbal and tonal information of the sequence was correct (same syllables, same pitches), (b) only the tonal, or (c) only the verbal sequence was correct, or (d) neither tonal nor verbal sequence were correct. Incongruency was obtained by exchanging the positions of two elements, that is, either of two syllables, or of two pitches, or of two pitch-syllable elements (see right of Figure 1). Each probe sequence type occurred equiprobably in the four conditions (V R, V S, T R, T S).

Participants had two response buttons (correct/incorrect) which they pressed with their left and right index finger. Key assignment was counterbalanced across participants. The fMRI experiment comprised 36 trials in each of the six conditions (T S, T R, V S, V R, N S, N 0), resulting in a total of 216 trials (with 216 different initial stimulus sequences), corresponding to a duration of approximately 50 minutes. During the experiment, trials of all six conditions were pseudo-randomly intermixed.

Participants were trained in a separate session of about one hour duration on a separate day within the week prior to the fMRI measurement. In the training session they performed all tasks both covertly and overtly. This enabled us to control that participants actually rehearsed only either the pitches or the syllables in the rehearsal conditions, and that they actually sang the children's song during the suppression condition. Moreover, it allowed us to collect behavioural data for overt rehearsal and suppression, and to compare these data with the behavioural data obtained in the fMRI session.

fMRI Scanning Procedure. Scanning was performed on a 3-T scanner (Medspec 30/100, Bruker, Ettlingen). Prior to the functional recordings, anatomical slices were acquired. The anatomical slices had the same geometric orientation as the functional slices. Before each functional session, a high-resolution anatomical reference data set (T1-weighted) was acquired for each participant, which was standardised to the Talairach stereotactic space (Talairach & Tournoux, 1988). A spin-echo EPI sequence was used with a TE of 75 ms, a TR of 2000 ms, and an acquisition bandwidth of 100 kHz. Acquisition of the slices was arranged uniformly within the TR interval. The matrix acquired was 64x64 with a FOV of 19.2 cm, resulting in an in-plane resolution of 3x3 mm. Slice thickness was 5 mm with an interslice gap of 1 mm (14 slices were acquired, nine above the AC-PC plane). In the present study, we did not choose a sparse temporal scanning design because our primary interest was not to investigate perceptual mechanisms within the auditory cortex (and the larger number of acquisitions may increase the signal-to-noise ratio in non-auditory regions). However, we currently investigate whether the continuous scanning interferes more strongly with the maintenance of tonal than with the maintenance of verbal information during articulatory suppression (unpublished data).

fMRI Data Analysis. fMRI data were processed using the software package LIPSIA (Lohmann et al., 2001). Functional data were corrected for motion using a matching metric based on linear correlation. To correct for the temporal offset between the slices acquired in one scan, a cubic-spline-interpolation was applied. A temporal high-pass filter with a cutoff frequency of 1/72 Hz was used for baseline correction of the signal and a spatial gaussian filter with 5.65 mm FWHM was applied. Functional data were linearly registered with the Talairach stereotactic coordinate system (Talairach and Tournoux, 1988). The rotational and translational parameters were subsequently transformed by linear scaling to the standard size. The resulting parameters were then used to transform the functional slices using trilinear interpolation, so that the resulting functional slices were aligned with the Talairach coordinate system.

The statistical evaluation was based on a least-squares estimation using the general linear model for serially autocorrelated observations (see also Friston, 1994; Worsley and Friston, 1995; Aguirre et al., 1997; Zarahn et al., 1997). The design matrix was generated using a box-car function. The design matrix, the acquired data, and the error term were convolved with a Gaussian kernel of 4 s (to deal with the temporal autocorrelation). Subsequently, contrast-images were calculated for each participant, and entered into a second-level random effects analysis. One-sample *t*-tests were performed to evaluate whether observed differences were significantly different from zero (*t*-values were transformed into *z*-values). The results were corrected for multiple comparisons using cluster-size and cluster-value thresholds obtained by Monte-Carlo simulations using a significance level of $p < 0.05$ (clusters in the resulting maps were obtained using a *z*-value threshold of 2.58).

Region of interest (ROI) analysis. Cortical areas that were significantly activated in the SPMs in either hemisphere, and in either the verbal or the tonal rehearsal condition, were subjected to a further post hoc analysis. It was tested whether the activation strength in regions of interest (ROIs) differed between hemispheres and/or conditions (for comparisons between hemispheres, or conditions, some areas were investigated with ROI analyses, even if those areas were not significantly activated in the SPMs with the applied statistical thresholds, see Table 1). For each subject, five ROIs were defined in each hemisphere and for each condition as single voxels. These ROIs were: (1) ventrolateral premotor cortex (vlPMC), (2) dorsolateral premotor cortex (dlPMC), (3) supramarginal gyrus / intraparietal sulcus (SMG/IPS), (4) planum temporale, and (5) the anterior superior insula. An additional ROI was defined as voxel in the pre-SMA. The ROI coordinates were determined separately for each subject (using for each subject the individual z-maps and the individual high-resolution anatomical scan). Such individually adjusted ROIs were computed because of the inter-individual variability of brain morphology, thus obtaining as high an accuracy of the statistical comparisons as possible (see Figure 3D for illustration of individual ROIs for verbal rehearsal). For the determination of ROI coordinates, SPMs were scaled to 1 x 1 x 1 mm using trilinear interpolation. The coordinate of each ROI was defined as pixel with the highest *z*-value in the interpolated single-subject SPM within a search radius of 9 mm around the local signal maximum in the group contrast, but within the anatomical boundaries of the respective structure (regardless of the statistical significance of the z-value; coordinates were determined by S.K. and T.F.). These coordinates were local maxima in approximately 95 % of all cases. Then, for each subject contrast values were computed for each contrast for the voxel containing the respective coordinate. These mean values subsequently entered repeated measures ANOVAs with factors condition (verbal rehearsal, tonal rehearsal) and hemisphere (Bosch, 2000). In addition to comparing the hemodynamic responses in the ROIs, coordinates of ROIs were compared between verbal and tonal rehearsal to test for possible differences in the topography of activations between these two conditions. Therefore, *x*-, *y*-, and *z*-coordinates were compared by paired two-sided *t*-tests.

The analogous procedure was applied for the maintenance of verbal, and the maintenance of tonal information during articulatory suppression with the following ROIs: (1) intermediate frontal sulcus, (2) IFG/pars triangularis, (3) vlPMC, (4) anterior superior insula, and (5) pre-SMA. To test for hemispheric differences, this procedure was also applied for the singing condition with the following ROIs: dlPMC, Rolandic operculum, planum temporale / supramarginal gyrus, IPL / angular gyrus, precuneus, and posterior cingulate cortex (PCC).

RESULTS

Behavioural data. The behavioural data for both verbal (syllable) and tonal (pitch) tasks are summarized in Figure 2. During the verbal rehearsal, participants had on average 97.25% (*SEM* = 0.78%) correct responses. Memory performance in the verbal task clearly dropped during articulatory suppression (87.08%, *SEM* = 2.60%). During the tonal rehearsal, participants had on average 63.83% (*SEM* = 2.82%) correct responses. Like in the verbal task, performance in the tonal task was less accurate under articulatory suppression (60.08%, *SEM* = 2.82%). To fulfil the requirements of normal distribution and equality of variances for an ANOVA, behavioural data for both verbal and tonal tasks were transformed with 2 * arcsin(sqrt(x)) (and 1-(1/2n) for x = 100% correct responses, n being the number of trials, see Kirk, 1995). A subsequent ANOVA with factors memory-type (verbal, tonal) and suppression (with, without articulatory suppression) on the hit rates showed a main effect of memory-type (*F*(1,11) = 131.78, *p* < .0001, reflecting that participants' performance was better in the verbal than in the tonal tasks), a main effect of suppression (*F*(1,11) = 15.82, *p* = .002, reflecting that performance was better during rehearsal than during articulatory suppression), and a two-way interaction (*F*(1,11) = 17.37, *p* = .002, reflecting that the effect of articulatory suppression on the memory task was stronger during the verbal than during the tonal task). Importantly, performance dropped significantly during suppression (compared to rehearsal) in both the verbal (*t*(11) = 4.27, *p* < .001) and the tonal task (*t*(11) = 2.22, *p* < .05), providing assurance that participants actually performed the articulatory suppression during both tasks. In the control task in which subjects only had to sing subvocally, without memorizing pitches or syllables, they classified 98.08% (*SEM* = 0.79) of the probe stimuli correctly as played forward/backward, and 99% (*SEM* = 0.56) in the control task in which they neither sang, nor memorized the stimuli.

During the rehearsal tasks, and during the verbal WM task under suppression, performance during the fMRI session was similar to the performance during the training session in which both rehearsal and suppression was also performed overtly (and could, thus, be controlled by the experimenter): Correct responses for verbal rehearsal were 96.64%, for maintenance of verbal information under articulatory suppression 83.35%, and for tonal rehearsal 70.04%. Paired *t*-tests showed that the differences between training and fMRI session were statistically not significant (verbal rehearsal: *p* > .7, maintenance of verbal information under suppression: *p* > .2, tonal rehearsal: *p* > .1). This provides some assurance that participants followed the instructions correctly. For the maintenance of tonal information during articulatory suppression, performance was significantly better during the training session (69.09% correct responses, $p < .05$), perhaps because the scanner noise made the tonal task more difficult. However, the fact that performance dropped during the fMRI experiment corroborates that the participants followed the instructions correctly.

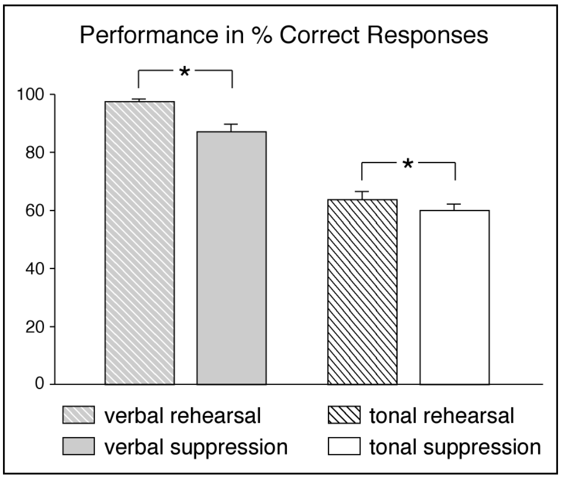

Figure 2: Behavioural data of verbal and tonal WM during rehearsal (shaded bars), and of verbal and tonal WM under simultaneous articulatory suppression (non-shaded bars). Participants performed better in the verbal than in the tonal conditions. Note the significant drop in performance (compared to the rehearsal conditions) during maintenance of both verbal and tonal information under articulatory suppression.

fMRI data: Rehearsal conditions. Table 1 summarizes activations elicited by the verbal and the tonal rehearsal (both contrasted to non-rehearsal, see also Figure 3A & B). The topography of local maxima was remarkably similar for both rehearsal conditions: Both the verbal and the tonal rehearsal activated a cortical network comprising (a) ventrolateral premotor cortex (vlPMC, this activation extended along the precentral sulcus into the posterior wall of the pars opercularis / Broca's area), (b) dorsolateral premotor cortex (dlPMC), (c) the intraparietal sulcus (IPS) extending into the supramarginal gyrus (SMG), (d) the planum temporale, (e) the anterior superior insula, (f) the pars triangularis of the IFG (BA 45/46), although during verbal rehearsal only in the right hemisphere), and (g) the pre-SMA [local maxima in the planum temporale lay within the probability region of 26-45 % for the planum temporale according to the probability maps of Westbury et al. (1999)]. This network clearly resembles the functional architecture of articulatory rehearsal reported in previous studies (see Introduction).

		left hemisphere					*right hemisphere*				
Anatomical structure	BA	Talairach coord. (SPM)	*z*-value (SPM)	mm^3	Talairach coord. (ROI)	p-value (ROI)	Talairach coord. (SPM)	*z*-value (SPM)	mm^3	Talairach coord. (ROI)	*p*-value (ROI)
Verbal rehearsal											
vlPMC	6	-44 1 24	4.6	2511	-46 3 22	.005				50 9 22	.007
dlPMC	4/6	-50 -8 42	8.27	1350	-48 -6 44	.0001	49 -8 39	5.35	378	49 -6 41	.0001
IPS/SMG	40	-38 -38 39	4.83	1161	-41 -37 40	.0002				41 -37 43	.0004
planum temporale	22	-44 -38 21	6.88	1269	-47 -40 22	.0001				52 -34 23	.0001
ant. sup. Insula					-32 15 4	n.s.				34 15 0	n.s.
IFG/pars trinangularis	45/46				-45 29 9	n.s.				45 32 8	.005
pre-SMA	6	-5 3 54	5.11	243	-6 6 55	.005					
subcentral gyrus[1]	43	-59 -5 15	5.94								
Putamen		-17 0 15	4.10	2511							
Caudate nucleus							16 10 15	4.01	594		
Thalamus[2]		-17 -18 15	4.25								
Cerebellum							25 -62 -15	5.68	1269		
Tonal rehearsal											
vlPMC	6	-50 4 24	12.97	8883	-49 6 22	.0001	49 7 21	9.43	2754	47 7 20	.0001
dlPMC	4/6	-47 -8 42	9.20	1080	-47 -5 43	.0001				50 -7 42	.0001
IPS/SMG	40	-47 -35 39	10.95	5238	-41 -38 40	.0001	34 -38 42	10.51	2403	40 -39 44	.0002
planum temporale[3]	22	-47 -42 24	8.07		-48 -41 26	.0001				51 -34 23	.0004
ant. sup. Insula [4]		-32 19 3	9.83		-31 18 5	.02				35 16 0	n.s.
IFG/pars trinangularis	45/46	-44 31 9	7.67	486	-43 30 10	.005				45 33 9	.005
pre-SMA	6	-5 7 54	10.85	3132	-5 7 55	.0005					
Caudate nucleus							16 9 15	4.01	189		
Pallidum		-17 -3 3	7.24	189							
Thalamus		-14 -14 15	7.31	918			13 -15 15	6.27	135		
Cerebellum		-29 -62 -18	9.62	864			31 -56 -18	8.47	1863		

[1] The cluster in the left vlPMC had another local maximum in the subcentral gyrus
[2] The cluster in the left striatum had another local maximum in the Thalamus
[3] The cluster in the left IPS/SMG had another local maximum in the planum temporale
[4] The cluster in the left vlPMC had another local maximum in the anterior superior insula

Table 1: Activations elicited during rehearsal conditions (contrasted to the control condition in which subjects did neither sing nor memorize). The table shows the results of the cluster analysis of statistical parametric maps ($p < 0.05$ corrected for multiple comparisons) and ROI analyses (see *Methods* for details).

The conjunction analysis showed that vlPMC, dlPMC, and SMG/IPS (all bilaterally), and the left planum temporale were significantly activated during both verbal and tonal rehearsal (Figure 3C). Moreover, activations for both conditions were also indicated in the pre-SMA, the cerebellum bilaterally, the left Rolandic operculum, putamen, pallidum, and thalamus, as well as the right caudate nucleus (not shown in Figure 3C).

For the verbal rehearsal an additional activation was indicated within the subcentral gyrus (Rolandic operculum, BA43). Such activation was not yielded for the tonal rehearsal in the corrected SPMs. However, a local maximum within this structure was also indicated for the tonal rehearsal in the uncorrected SPMs ($z = 5.45$; coordinate of this local maximum was -52, -14, 14), strongly suggesting that the Rolandic operculum was not only activated during the verbal, but also during the tonal rehearsal.

To investigate hemispheric differences, and to compare both activation patterns in more detail, ROI analyses were performed using individually adjusted ROIs (see Table 1 for results, see Figure 3D for illustration of individual ROIs): For each participant and each analyzed structure, a ROI-coordinate was determined as local maximum of activation within the anatomical boundaries of the respective structure (vlPMC, dlPMC, planum temporale, SMG/IPS, and pre-SMA, see *Methods* for details).

In a first step, we investigated if these ROI coordinates differed between verbal and tonal rehearsal. Therefore, individual ROI coordinates were grand-averaged separately for each structure (grand-averaged ROI-coordinates are provided in Table 1). In each of the analyzed structures (except the planum temporale) grand-averaged coordinates of local maxima of the verbal rehearsal were located within a 3 mm range of the respective coordinates of the tonal rehearsal (in the planum temporale, grand-averaged coordinates were within a 4 mm range). That is, given the spatial resolution of our fMRI data, the local maxima of ROI coordinates were virtually identical for both verbal and tonal rehearsal (located within the same, or the directly adjacent voxel).

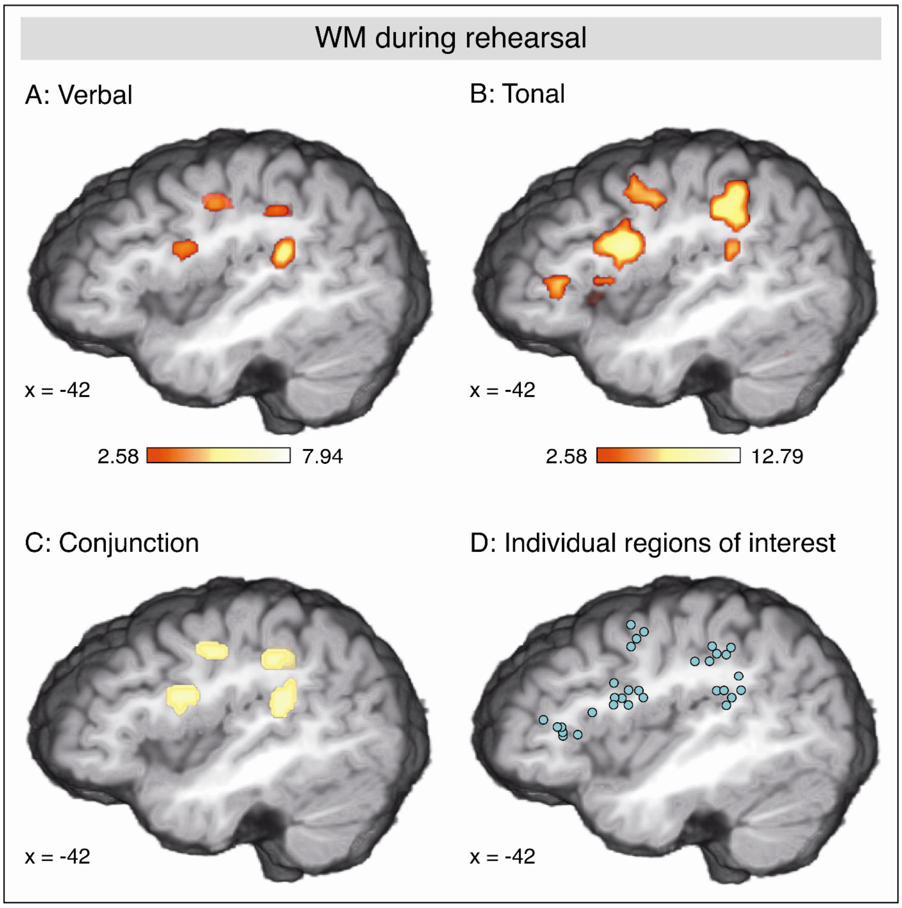

Figure 3: Activations during verbal (**A**) and tonal (**B**) rehearsal (contrasted to the control condition in which subjects did neither sing nor memorize; $p < 0.05$ corrected for multiple comparisons). Both tasks activated a network comprising the ventrolateral premotor cortex (vlPMC), the dorsal precentral gyrus, the intraparietal sulcus (IPS) extending into the supramarginal gyrus (SMG), and the planum temporale (p.t.). In the left hemisphere, the pars triangularis of the IFG was activated only during the tonal rehearsal. C shows areas that were significantly activated during both verbal and tonal rehearsal. D: Illustration of individually adjusted ROIs (for the verbal rehearsal condition). In each cluster, each circle represents the ROI coordinate of one participant.

These observations were confirmed by statistical analyses: Paired *t*-tests on the *x*-, *y*-, and *z*-coordinates of individual ROI coordinates were computed for each structure to test if the coordinates of activations during tonal rehearsal differed from coordinates of activations during verbal rehearsal. These *t*-tests did not indicate any difference between verbal and tonal rehearsal (*p* was between .1 and .2 in four tests, between .21 and .89 in 33 tests, and > .9 in two tests). This indicates that the functional architecture of verbal and tonal rehearsal does not differ, at least when applying the task used in the present study.

To test differences in activation strength between conditions, and lateralization of activations, ANOVAs with factors condition (tonal rehearsal, verbal rehearsal) and hemisphere were carried out for each ROI, indicating significant effects of hemisphere for the vlPMC (*p* < .05), and a marginally significant effect of hemisphere for the SMG/IPS (p < .07) as well as for the planum temporale (*p* < .07). Significant effects of condition were indicated for BA 45/46 (*p* < .02), and for the vlPMC (p < .05). A *t*-test comparing left and right BA46 for the tonal condition only indicated a significant difference between hemispheres (p < .05).

fMRI data: Suppression conditions. Figure 4 (A & B) shows the activations during verbal and tonal WM under articulatory suppression (i.e., singing the children's song while maintaining the pitches or the syllables in WM) contrasted to the control condition (singing the children's song without keeping the pitches or syllables in memory). Significant activations were observed for both verbal and tonal conditions within the left vlPMC (extending into the pars opercularis / Broca's area), the anterior insula, the right cerebellum and the right striatum (see Figure 4, and Table 2). Moreover, activations were present in the IFG (pars triangularis, BA 45/46) during the tonal condition, and during both tonal and verbal conditions in the inferior frontal sulcus (IFS, see also Table 2). The latter activation extended anteriorly along the upper bank of the IFS into the frontomarginal / anterior intermediate frontal sulcus (see inset in Figure 4). The conjunction analysis showed that, in the left hemisphere, vlPMC, and anterior prefrontal areas were significantly activated during maintenance of both verbal and tonal information in the face of simultaneous suppression (Figure 4C). Moreover, activations for both conditions were observed in the anterior insula bilaterally, as well as in the right putamen and the right cerebellum.

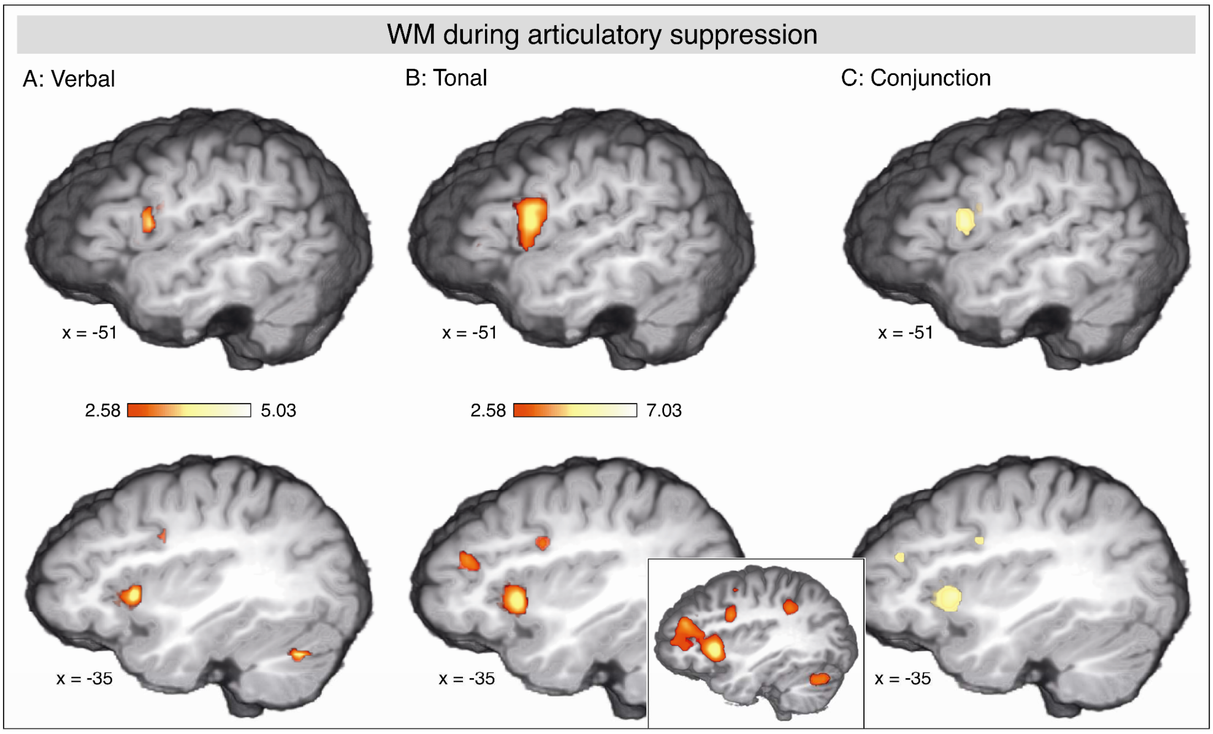

Figure 4: Activations during maintenance of verbal (**A**) and maintenance of tonal (**B**) information under articulatory suppression (contrasted to the control condition in which subjects covertly sung, but did not memorize; *p* < 0.05 corrected for multiple comparisons). During both verbal and tonal conditions, activations were observed in the vlPMC (extending into the pars opercularis / Broca's area), the anterior insula, the right cerebellum and the right ventral striatum (not shown). Additional activations were indicated in the pars triangularis, and in the inferior frontal sulcus. The inset in **B** shows that the latter activation extended anteriorly along the upper bank of the IFS into the frontomarginal / intermediate frontal sulcus (*p* < 0.05 uncorrected). C shows areas that were significantly activated during both conditions.

In contrast to the rehearsal condition, no significant activations were indicated for the dlPMC, or the planum temporale (the IPL was activated only during the tonal condition, and only in the left hemisphere). Because the absence of significant activations in the SPMs does not indicate that these structures were completely inactive, each coordinate of the network observed under verbal and tonal rehearsal was examined by searching for the nearest local maximum in the *z*-maps of the suppression contrasts (only activations with a *p*-value of at least .05 uncorrected were regarded as local maxima). In all structures of the left hemisphere, except the planum temporale, local maxima were found within the same, or the adjacent, voxel as in the rehearsal conditions. These findings were supported by ROI analyses (using the coordinates determined for the rehearsal conditions): All mentioned structures (vlPMC, dlPMC, SMG/IPS, and insula), but not the planum temporale, were also significantly activated (all $p < .05$) during the suppression conditions. That is, the activity of the network observed under articulatory rehearsal was not completely abolished, although strongly reduced during articulatory suppression.

In addition to this ROI analysis (which used ROI coordinates obtained for the rehearsal conditions), we also obtained the individual coordinates of activations during the suppression conditions (grand-averaged ROI-coordinates are provided in Table 2). As in the rehearsal conditions, in each of the analyzed structures (except the vlPMC) grand-averaged coordinates of local maxima of the verbal and the tonal conditions were located within the same, or the directly adjacent voxel. In the vlPMC, the ROI coordinates differed between the verbal and the tonal condition (with regards to *x*-, *y*-, and *z*-coordinates, $p < .05$ in all three paired *t*-tests). No such differences between conditions were indicated for any other structure (neither in *x*-, *y*-, nor *z*-direction, $p > .2$ in each test). To test for differences in activation strength between conditions, and for lateralization of activations, ANOVAs with factors condition (maintenance of tonal, and maintenance of verbal information, both during articulatory suppression) and hemisphere were carried out for each ROI, but no main effects or interactions were indicated.

Figure 5 shows activations of the covert singing (contrast: singing vs. not singing, without memorizing pitches or syllables in both conditions, see also Table 3). Marked activations were found within the planum temporale bilaterally (in the left hemisphere extending into the supramarginal gyrus), the Rolandic operculum bilaterally, and the dlPMC bilaterally. Notably, in contrast to the rehearsal and suppression conditions, activations within the vlPMC or Broca's area were not significant with the applied statistical threshold.

		left hemisphere									*right hemisphere*								
Anatomical structure	BA	alairach oord. (SPM)			*z*-value (SPM)	nm^3	Talairach coord. (ROI)			)-value (ROI)	Talairach Coord. (SPM)			*z*-value (SPM)	mm^3	Talairach coord. (ROI)			*p*-value (ROI)
Verbal WM (during suppression)																			
intermediate frontal sulcus		-35	40	24	3.13	111	-34	38	23	.05						34	40	11	n.s.
IFG/pars trinangularis	45/46						-44	27	3	n.s.						48	30	1	n.s.
vlPMC	6	-53	7	15	4.25	648	-48	8	16	.01						47	5	18	.05
IPS/SMG	40						-44	-36	41	.05						40	-40	41	.08
ant. sup. Insula		-29	19	3	5.09	1269	-29	20	3	.0005	37	16	3	3.98	729	33	16	2	.001
Pre-SMA	6	-5	16	51	5.05	2511	-5	14	52	.0001									
Putamen											22	16	-3	4.03	270				
Pallidum		-14	-6	0	3.79	162													
Cerebellum		-38	-65	-24	4.75	432					25	-59	-18	3.82	891				
Tonal WM (during suppression)																			
intermediate frontal sulcus		-35	40	21	3.99	648	-33	39	21	.05						35	43	14	.08
IFG/pars trinangularis	45/46	-41	37	6	3.86	270	-44	29	3	.05						47	30	2	n.s.
vlPMC[1]	6	-44	3	24	5.42		-44	3	23	.005						48	5	24	.05
IPS/SMG	40	-50	-35	48	3.76	243	-46	-37	44	.05						41	-41	44	.05
ant. sup. Insula		-29	19	6	7.14	5751	-28	19	3	.001	34	19	6	4.52	3159	31	17	3	.01
pre-SMA	6	-5	25	45	6.36	5940	-3	14	51	.0001									
Cerebellum											31	-56	-24	4.05	216				

[1] The cluster in the insula had another local maximum in the vlPMC

Table 2: Activations elicited during maintenance of verbal, and maintenance of tonal information under articulatory suppression (contrasted to the control condition in which subjects covertly sang, but did not memorize). The table shows the results of the cluster analysis of statistical parametric maps ($p < 0.05$ corrected for multiple comparisons) and ROI analyses.

DISCUSSION

Rehearsal. During the verbal rehearsal, a neural network including the vlPMC and dlPMC, the anterior insula, the SMG/IPS, the planum temporale, the IFG, pre-SMA and the cerebellum was activated. This network has been described in previous studies on verbal WM with auditory (Hickok et al, 2003), and visual stimuli (with the exception of the planum temporale; Awh et al., 1996; Chen & Desmond, 2005; Gruber, 2001; Gruber & von Cramon, 2001; Gruber & von Cramon, 2003; Kirschen et al., 2005; Paulesu et al., 1993). Importantly, virtually the identical network as during verbal rehearsal was also found to be activated during the tonal rehearsal: the coordinates of the above mentioned activations did not differ within subjects between the verbal and the tonal rehearsal, and the conjunction analysis showed that the mentioned structures (in concert with the left putamen, pallidum, and thalamus, as well as the right caudate nucleus) were significantly activated in both conditions. Thus, the present data show a remarkable overlap of (neo)cortical, subcortical, and cerebellar neuronal resources underlying the rehearsal of verbal and tonal stimuli, indicating that phonological rehearsal mechanisms are less specialized for language than usually believed. This finding is in agreement with previous studies on tonal WM (Gaab et al., 2003; Zatorre et al., 1994; Hickok et al, 2003) which reported activations for tonal WM that were similar to those observed in studies on verbal WM (see above). Our data are also in line with previous findings from Hickock et al. (2003), who observed activations of the vlPMC, the IPS/SMG, the planum temporale (referred to as *area Spt* by the authors), and the dlPMC for the rehearsal of both melodies and sentences.

Differences between conditions were activations of the triangular part of the left inferior frontal gyrus, and of the left anterior superior insula during the tonal, but not during the verbal rehearsal task. However, it is unlikely that these regions play a role only for tonal WM: Both the left anterior insula (Bamiou et al., 2003; Paulesu et al., 1993; Chein et al., 2002) as well as the triangular part of the left inferior frontal gyrus (e.g., Cabeza & Nyberg, 2000) have also been reported to be involved in verbal WM tasks (as well as in other tasks such as speech production, and action observation, e.g. Augustine, 1996; Binkofski et al., 1999; Buccino et al., 2001). During WM tasks, the latter region appears to become particularly involved when strategic processes come into play, such as organizing of WM contents into higher level chunks (see Bor et al., 2003; that study used a visual-spatial task). Because the tonal WM task was more difficult than the verbal task (as reflected in the behavioural data), it is likely that participants engaged strategic processes (especially chunking of the pitches of a sequence into melodic segments) more strongly during tonal rehearsal than during verbal rehearsal, and that this engagement may be reflected by the activation of BA 45/46 during the tonal rehearsal. The greater difficulty of the tonal task presumably also explains that activations were stronger during the tonal than during the verbal rehearsal. Taken together, the present data thus indicate that rehearsal of tonal information (i.e., a “tonal loop”) relies on neural resources that strongly overlap in their topography with those involved in the rehearsal of verbal information.

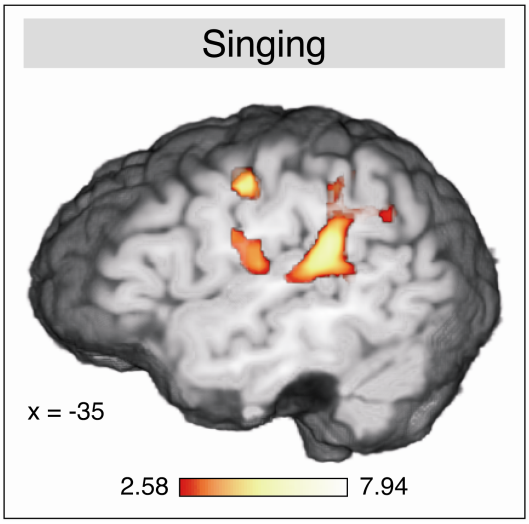

Figure 5: Activations elicited during covert singing (contrasted to the control condition in which subjects did not sing; $p < 0.05$ corrected for multiple comparisons).

It is unlikely that the similar topography of activations for verbal (syllable) and tonal (pitch) rehearsal is simply due to non-compliant behaviour on the part of the subjects (i.e., simultaneous rehearsal of both syllables and pitches): *First*, the behavioural data recorded during the training-session (in which participants rehearsed overtly, allowing the experimenter to control that they were rehearsing either only the tones, or only the syllables, but not both simultaneously) were very similar to those recorded during the covert rehearsal of the fMRI session. *Second*, after a little training, it is considerably easier, and more convenient, to rehearse either the tones or the syllables (instead of both, this was also mentioned by our participants when asked after the experiment). *Third*, activations were in some structures significantly stronger during the tonal rehearsal (compared to the verbal rehearsal), which is not to be expected if participants performed the same rehearsal (i.e., simultaneous rehearsal of syllables and pitches) in both conditions. Finally, fMRI research experience tells us that subjects are generally highly compliant in covert tasks (see also Callan et al., 2006).

(Non)specificity of activations during rehearsal. During subvocal articulatory rehearsal, strong activations of the vlPMC, extending anteriorly into the precentral sulcus, and Broca's area were observed. Such activations were not observed during the subvocal singing, suggesting that vlPMC/Broca's area plays a more specific role for verbal and tonal WM. The present data, thus, point to the particular importance of vlPMC as an active rehearsal component (which is a substantial part of the articulatory loop). However, it should also be noted that the vlPMC is not only involved in WM functions: Previous studies have also shown involvement of this region in a number of other functions such as action planning and -understanding, serial prediction, and analysis as well as recognition of sequential information (Buccino et al., 2001; Conway and Christiansen, 2001; Huettel et al., 2002; Rizzolatti & Craighero, 2004; Schubotz and von Cramon, 2002; Meyer & Jancke, 2006; see also below).

Although the SMG and the IPS were also active during the singing condition, activations of these areas appeared to be considerably stronger during the rehearsal conditions. Because these structures have previously been reported to play an important role for WM (e.g., Awh et al., 1996; Crottaz-Herbette et al., 2004; Gruber, 2001; Gruber & vonCramon, 2003), and because singing also involves WM operations, it is likely that these areas serve WM processes, rather than simply articulatory processes. For example, following a suggestion by Cohen et al. (1997), these inferior parietal areas may store phonological long-term information that may be actively accessed via item-specific functional connections to the anterior prefrontal cortex, which has been shown to play a major role both in verbal WM (Gruber, 2001; Gruber & von Cramon, 2003) and in memory retrieval (see, for example, Buckner & Koutstaal, 1998). Nevertheless, we already noted in the Introduction that regions along the IPS are also involved in attentional mechanisms (and in a number of other functions such as spatial localization, reaching and grasping, as well as task switching, for overviews see Cabeza & Nyberg, 2000; Culham & Kanwisher, 2001; Corbetta & Shulman, 2002). Hence, further studies are needed to clarify the particular role that this region plays for WM.

By contrast, the planum temporale was not only activated during the subvocal rehearsal, but activated even more strongly during the covert singing condition. This suggests that the planum temporale plays a role for mechanisms that are not directly dependent on WM processes, such as the formation of auditory images during rehearsal or singing (Halpern & Zatorre, 1999), transformation of such images into motor codes (Warren et al., 2005; Hickock et al., 2003; Buchbaum et al., 2005; Callan et al., 2006), segregation and analysis of the spectrotemporal structure of sounds (Griffiths & Warren, 2002; Jäncke et al., 2002; Binder et al., 2000), as well as matching of spectrotemporal patterns with learned spectrotemporal representations (Griffiths & Warren, 2002).

Like the planum temporale, the dorsal precentral gyrus was not only activated during rehearsal, but also during subvocal singing. The coordinates of the dlPMC activations were virtually identical between the singing and rehearsal conditions, and also highly similar to the coordinates reported for monotonic vocalizations of tones reported in a previous study (Brown et al., 2004; in that study, the coordinate reported for monotonic vocalization was x = -48, y = -10, z = 44). This indicates that this region of the dorsolateral premotor cortex serves articulatory processes independent of WM operations.

Suppression. Activations in the dlPMC, the planum temporale, and the IPL were considerably smaller during the suppression conditions (i.e., singing with the additional task of remembering the syllables and the pitches) compared to the rehearsal conditions. This indicates that the articulatory suppression impaired the phonological loop, which is also reflected in the behavioural data. However, activations within these regions were not completely abolished. The possible reasons for the residual activations of these areas during suppression are discussed in the next section.

Interestingly, particularly during the maintenance of tonal information under articulatory suppression, activations within the IFG/IFS extended into the frontomarginal sulcus/anterior intermediate frontal sulcus. This further supports previous findings suggesting that these anterior prefrontal areas constitute an important component of a (bilateral) prefrontal-parietal network that becomes activated whenever the speech-based rehearsal mechanism is not available, or not sufficient, to solve a memory task by itself (Gruber, 2001; Gruber & von Cramon, 2001, 2003; Gruber & Goschke, 2004; Gruber et al., 2005; Gruber et al., 2007). Therefore, these areas might contain additional storage components of WM that are activated when auditory information cannot be rehearsed (Gruber & von Cramon, 2003). Note that the latter study (Gruber & von Cramon, 2003) also showed that these regions are selectively activated in a (phonological) WM task under articulatory suppression, but not under similar conditions of conflict in the visuospatial domain (i.e. during visuospatial WM under visuospatial suppression), providing evidence that these regions are not simply involved in general executive control that support other WM areas in situations of conflict, but rather subserve domain-specific processes related to phonological WM. Because these areas were activated during both verbal and tonal WM under suppression conditions, our data indicate that this additional WM component is important for the storage of both verbal and tonal information.

Rehearsal and Suppression: Sensorimotor codes. Previous work has suggested that sensorimotor processes may assist with the representation and manipulation of information, and that sensorimotor coding plays an important role for WM processes (for a review see Wilson, 2001). Sensorimotor codes provide resources for the representation and maintenance of information (in the present study verbal and tonal information), and it is highly plausible that such resources were used by the participants to perform the WM tasks. This assumption is supported by the strong activation of lateral premotor areas along with parietal areas, cerebellar, and subcortical regions during the rehearsal conditions (and, although to a lesser extent, also during the suppression conditions). Numerous neurons in these regions are also involved in cortico-basal ganglia-thalamo-cortical and cerebellar loops that serve voluntary motor control, and contribute to the programming, initiation, and execution of movements (Parent & Hazrati, 1995; Leblois et al., 2006; Hoover & Strick, 1999; Middleton & Strick, 2000).

Also note that motor actions are not only coded by premotor, but also by parietal areas and that, in addition, parietal areas translate sensory input into information appropriate for action, and provide representations of these actions with specific sensory information (e.g., Fogassi et al., 2005; Fogassi & Luppino, 2005). In the present study, the parietal (SMG/IPS), and ventrolateral as well as dorsolateral premotor areas observed to be active during the rehearsal tasks (along with subcortical and cerebellar structures) might thus represent neural circuits involved in the formation and maintenance of sensorimotor codes serving the rehearsal of the tonal and verbal information. Because sensorimotor coding is involved in a number of different tasks (such as observing, performing, or recognizing actions), parts of the network observed in the present study (particularly the premotor and parietal regions) have also been reported in a number of previous studies that did not focus on WM (for an overview see Rizzolatti & Craighero, 2004, see also Janata et al., 2002). Likewise, one reason for the strong overlap of neuronal networks involved in verbal and tonal WM functions is presumably that WM for phonemes and for pitches relies to a considerable amount on sensorimotor-related circuits which are similar for speech and song (see also Callan et al., 2006; that study showed a remarkable overlap of the brain structures involved in covert singing and covert speech, among them dlPMC and the planum temporale).

The residual activations of these areas during suppression are possibly due to the formation of motor representations during the presentation of stimuli in the suppression trials, which were not instantly erased with the onset of the suppression, but which probably decayed during the suppression, and were, thus, still residually observable. That is, it appears likely that the verbal and tonal information was encoded in sensorimotor representations, that these codes were held active during rehearsal, but decayed during suppression. However, on the other hand it cannot be excluded that the activations of SMG/IPS, and of vlPMC during the suppression conditions were simply due to erroneous rehearsal in some trials by some participants. Nevertheless, the strong decrease of performance during the suppression condition (as reflected in the behavioural data) indicates that participants mainly followed the instructions correctly. During articulatory suppression, the local maximum of activation in the vlPMC appeared to differ between tonal and verbal condition (as indicated by the ROI coordinates). However, because no such difference was found in the rehearsal conditions, and because the conjunction analysis showed a clear overlap of both tonal and verbal WM under suppression, we suggest to await whether future studies can replicate this effect.

Singing. A nice additional finding was the activation of the Rolandic operculum during the singing condition (as well as during rehearsal). Similar activations have been reported in previous functional imaging studies on both overt and covert singing (Jeffries et al., 2003; Riecker et al., 2000; Wildgruber et al., 1996). The Rolandic operculum has been proposed to contain the representation of the larynx (and the pharynx), that is, of a vocal tract articulator crucially involved in the production of melody (Koelsch et al., 2006). The present results support this assumption, and highlight the importance of this area for the production of frequency-modulated vocal signals.

In conclusion, our data show that the topography of (neo)cortical, subcortical, and cerebellar WM components is strongly overlapping for the rehearsal of verbal and tonal information, as well as for the maintenance of such information during articulatory suppression. This indicates that the functional architecture of verbal and tonal WM is remarkably similar. Articulatory rehearsal of verbal and tonal information involved mainly motor-related areas (along with basal ganglia and thalamic nuclei, as well as the cerebellum), whereas maintenance of information during articulatory suppression additionally involved anterior prefrontal areas which might contain additional storage components of WM that are activated when auditory information cannot be rehearsed. The overlap of the neuronal networks underlying verbal and tonal WM, and the involvement of brain structures implicated in sensorimotor processing suggests that WM for phonemes and for pitches relies considerably on sensorimotor-related circuits which are similar (and partly identical) for speech and song. Because of such sensorimotor coding, some WM circuits are also overlapping with circuits involved in other cognitive tasks which do not involve WM, but require activity of sensorimotor-related processes.

References

Augustine, JR (1996). Circuitry and functional aspects of the insular lobe in primates including humans. Brain Research Reviews 22:229-244.

Awh E, Jonides J, Smith EE, Schumacher EH, Koeppe RA, Katz S. (1996): Dissociation of storage and rehearsal in verbal working memory: evidence from positron emission tomography. Psychol Sci 7(1):25-31.

Baddeley AD. (1992): Working memory. Science 255(5044):556-559.

Baddeley AD. 2003. Working memory: looking back and looking forward. Nat Rev Neurosci 4(10):829-839.

Baddeley AD, Hitch GJ. (1974) Working memory. In Bower GA editor. Recent Advances in Learning and Motivation (Vol. VIII, pp. 47-89). New York: Academic Press.

Bamiou DE, Musiek FE, Luxon LM. (2003) The insula (Island of Reil) and its role in auditory processing. Brain Res Rev 42(2):143-154.

Binder JR, Frost JA, Hammeke TA, Bellgowan PS, Springer JA, Kaufman JN, et al. (2000): Human temporal lobe activation by speech and nonspeech sounds. Cereb Cortex 10(5):512-528.

Binkofski F, Buccino G, Posse S, Seitz RJ, Rizzolatti G, Freund H (1999). A fronto-parietal circuit for object manipulation in man: evidence from an fMRI-study. Eur J Neurosci. 11(9):3276-3286.

Bor, D., Duncan, J., Wiseman, R. J., & Owen, A. M. (2003). Encoding strategies dissociate prefrontal activity from working memory demand. Neuron, 37(2), 361-367.

Bosch, V. 2000. Statistical analysis of multi-subject fmri data: The assessment of focal activations. *J. Magn. Reson. Imag.* **11:** 61–64.

Brown S, Martinez MJ, Hodges DA, Fox PT, Parsons LM. (2004). The song system of the human brain. Brain Res Cogn Brain Res. 20(3):363-375.

Buccino, G., Binkofski, F., Fink, G. R. , Fadiga, L. , Fogassi, L. , Gallese, V. , Seitz, R. J., Zilles, K., Rizzolatti, G. & Freund, H.-J. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. European Journal of Neuroscience 13 (2): 400–404.

Buchsbaum BR, Olsen RK, Koch PF, Kohn P, Kippenhan JS, Berman KF (2005). Reading, hearing, and the planum temporale. Neuroimage 24(2):444-454.

Buckner RL., Koutstaal W. (1998): Functional neuroimaging studies of encoding, priming, and explicit memory retrieval. Proc Natl Acad Sci USA 95(3):891-898.

Cabeza, R & Nyberg, L (2000). Imaging cognition II: An empirical review of 275 PET and fMRI studies. Journal of Cognitive Neuroscience, 12(1), 1-47.

Callan DE, Tsytsarev V, Hanakawa T., Callan AM, Katsuhara M, Fukuyama H, Turner B (2006). Song and speech: Brain regions involved with perception and covert production. NeuroImage 31:1327-1342.

Chan AS, Ho YC, Cheung MC. (1998): Music training improves verbal memory. Nature 396:128.

Chein JM, Fissell K, Jacobs S, Fiez JA. (2002): Functional heterogeneity within Broca's area during verbal working memory. Physiol. Behav 77:635-639.

Chen SH, Desmond JE. (2005): Cerebrocerebellar networks during articulatory rehearsal and verbal working memory tasks. Neuroimage 24(2):332-338.

Cohen JD, Perlstein WM, Braver TS, Nystrom LE, Noll DC, Jonides J, Smith EE. (1997): Temporal dynamics of brain activation during a working memory task. Nature 386:604-608.

Conway, C., Christiansen, M., 2001. Sequential learning in non-human primates. Trends Cogn Sci 5: 539–546.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. Nature Reviews Neuroscience, 3(3), 201-215.

Crottaz-Herbette S, Anagnoson RT, Menon V. (2004): Modality effects in verbal working memory: differential prefrontal and parietal responses to auditory and visual stimuli. Neuroimage 21(1):340-351.

Culham, J. C., & Kanwisher, N. G. (2001). Neuroimaging of cognitive functions in human parietal cortex. Current Opinion in Neurobiology, 11(2), 157-163.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading.Journal of Verbal Learning and Verbal Behavior, 19, 450-466.

Deutsch D. (1970): Tones and numbers: specificity of interference in immediate memory. Science 168(939):1604-1605.

Fiez, J. A., Raife, E. A., Balota, D. A., Schwarz, J. P., Raichle, M. E., & Petersen, S. E. (1996). A positron emission tomography study of the short-term maintenance of verbal information. Journal of Neuroscience 16(2): 808-822.

Fogassi L, Ferrari PF, Gesierich B, Rozzi S, Chersi F, Rizzolatti G (2005). Parietal lobe: from action organization to intention understanding. Science 308:644-5.

Fogassi L, Luppino G (2005). Motor functions of the parietal lobe. Curr Opin Neurobiol 15(6):626-631.

Gaab N, Gaser C, Zaehle T, Jancke L, Schlaug G. (2003): Functional anatomy of pitch memory--an fMRI study with sparse temporal sampling. Neuroimage 19(4):1417-1426.

Griffiths TD, Warren JD (2002). The planum temporale as a computational hub. Trends Neurosci. 25(7):348-353.

Gruber O. (2001): Effects of domain-specific interference on brain activation associated with verbal working memory task performance. Cereb Cortex 11:1047-1055.

Gruber O, von Cramon DY. (2003): The functional neuroanatomy of human working memory revisited. Evidence from 3-T fMRI studies using classical domain-specific interference tasks. Neuroimage 19(3):797-809.

Gruber O, Goschke T. (2004): Executive control emerging from dynamic interactions between brain systems mediating language, working memory and attentional processes. Acta Psych 115:105-121.

Gruber O, Gruber E, Falkai P. (2005):. Neural correlates of working memory deficits in schizophrenic patients. Ways to establish neurocognitive endophenotypes of psychiatric disorders. Radiologe 45:153-160.

Gruber O, Müller T, Falkai P. (2007): Dynamic interactions between brain systems underlying different components of verbal working memory. J Neural Trans, pub ahead of print.

Halpern AR, Zatorre RJ. (1999): When that tune runs through your head: a PET investigation of auditory imagery for familiar melodies. Cereb Cortex 9(7):697-704.

Henson, R. N., Burgess, N., & Frith, C. D. (2000). Recoding, storage, rehearsal and grouping in verbal shortterm memory: an fMRI study. Neuropsychologia, 38(4), 426-440.

Hickok G, Buchsbaum B, Humphries C, Muftuler T. (2003): Auditory-motor interaction revealed by fMRI: speech, music, and working memory in area. Spt J Cogn Neurosci 15(5):673-682.

Hoover JE, Strick PL (1999). The organization of cerebellar and basal ganglia outputs to primary motor cortex as revealed by retrograde transneuronal transport of herpes simplex virus type 1. J Neurosci. 19(4):1446-1463.

Huettel, S., Mack, P., McCarthy, G., 2002. Perceiving patterns in random series: dynamic processing of sequence in prefrontal cortex. Nat. Neurosci. 5, 485– 490.

Iwanaga M, Ito T. (2002): Disturbance effect of music on processing of verbal and spatial memories. Percept Mot Skills 94(3 Pt 2):1251-1258.

Janata P, Tillmann B, Bharucha JJ. (2002). Listening to polyphonic music recruits domain-general attention and working memory circuits. Cogn Affect Behav Neurosci. 2(2):121-140.

Jäncke L., Wüstenberg T., Scheich H., Heinze H.-J. (2002). Phonetic Perception and the Temporal Cortex. NeuroImage 15(4):733-746.

Jeffries KJ, Fritz JB, Braun AR (2003). Words in melody: an H(2)15O PET study of brain activation during singing and speaking. Neuroreport 14(5):749-754.

Jonides, J., Schumacher, E. H., Smith, E. E., Koeppe, R. A., Awh, E., Reuter-Lorenz, P. A., Marshuetz C & Willis CR (1998). The role of parietal cortex in verbal working memory. Journal of Neuroscience, 18(13): 5026-5034.

Kirk, R. E. (1995). Experimental design. ITP: Pacific Grove, Calif.

Kirschen MP, Chen SH, Schraedley-Desmond P, Desmond JE. (2005): Load- and practice-dependent increases in cerebro-cerebellar activation in verbal working memory: an fMRI study. Neuroimage 24(2):462-472.

Koelsch, S., Fritz, T., von Cramon, D.Y., Muller, K. & Friederici, A.D. (2006). Investigating emotion with music: an fMRI study. Hum. Brain Mapp., 27, 329-350.

Leblois A, Boraud T, Meissner W, Bergman H, Hansel D (2006). Competition between feedback loops underlies normal and pathological dynamics in the basal ganglia. J Neurosci 26(28):7317-7318.

Lohmann G, Muller K, Bosch V, Mentzel H, Hessler S, Chen L, et al. (2001): LIPSIA - a new software system for the evaluation of functional magnetic resonance images of the human brain. Comput Med Imaging Graph 25(6):449-457.

Meyer, M. & Jancke, L. (2006).Involvement of left and right frontal operculum in speech and non-speech perception and production.In: Y. Grodzinsky & K. Amunts (Eds.) Broca's region. New York: Oxford University Press, 218-241.

Middleton FA, Strick PL (2000) Basal ganglia and cerebellar loops:motor and cognitive circuits. Brain Res Brain Res Rev 31:236-250.

Oldfield, R.C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia, 9(1), 97-113.

Parent A, Hazrati LN (1995). Functional anatomy of the basal ganglia. I. The cortico-basal ganglia-thalamo-cortical loop. Brain Res Brain Res Rev. 20(1):91-127.

Paulesu E, Frith CD, Frackowiak RS. (1993): The neural correlates of the verbal component of working memory. Nature 362(6418):342-345.

Pechmann, T., & Mohr, G. (1992). Interference in memory for tonal pitch: implications for a working-memory model. Memory and Cognition, 20(3), 314-320.

Ravizza SM, Delgado MR, Chein JM, Becker JT, Fiez JA. (2004): Functional dissociations within the inferior parietal cortex in verbal working memory. Neuroimage 22(2):562-573.

Riecker A, Ackermann H, Wildgruber D, Dogil G, Grodd W (2000). Opposite hemispheric lateralization effects during speaking and singing at motor cortex, insula and cerebellum. Neuroreport 11(9):1997-2000.

Rizzolatti G, Craighero L. (2004). The mirror-neuron system. Annu Rev Neurosci. 27:169-92.

Schubotz, R.I., von Cramon, D.Y., 2002. Predicting perceptual events activates corresponding motor schemes in lateral premotor cortex: an fMRI study. NeuroImage 15, 787–796.

Salame P, Baddeley AD. (1989): Effects of background music on phonological short-term memory. Quarterly J Exp Psych 41(A):107-122.

Semal C, Demany L, Ueda K, Halle PA. (1996): Speech versus nonspeech in pitch memory. J Acoust Soc Am 100(2 Pt 1):1132-1140.

Talairach P, Tournoux J. (1988): A stereotactic coplanar atlas of the human brain. Stuttgart: Thieme.

Warren JE, Wise RJ, Warren JD. (2005): Sounds do-able: auditory-motor transformations and the posterior temporal plane. Trends Neurosci 28(12):636-43.

Westbury CF, Zatorre RJ, Evans AC (1999): Quantifying Variability in the Planum Temporale: A Probability Map. Cerebral Cortex 9:392-405.

Wildgruber D, Ackermann H, Klose U, Kardatzki B, Grodd W (1996). Functional lateralization of speech production at primary motor cortex: a fMRI study. Neuroreport 7:2791-2795.

Wilson M (2001): The case for sensorimotor coding in working memory. Psychonomic Bulletin & Review 8(1):44-57.

Zatorre, R. J., Evans, A. C., & Meyer, E. (1994). Neural mechanisms underlying melodic perception and memory for pitch. Journal of Neuroscience, 14(4), 1908-1919.

		Left hemisphere						*Right hemisphere*					
Anatomical structure	BA	Coordinate SPM)			*z*-value (SPM)	mm^3	*p*-value (ROI)	Coordinate (SPM)			*z*-value (SPM)	mm^3	*p*-value (ROI)
Singing													
dlPMC	4/6	-50	-8	42	6.27	810	.005	49	-8	39	6.03	459	.001
Rol. operc.	43	-59	-8	12	5.47	1134	.001						.01
p.t./SMG		-47	-38	21	7.38	5643	.0001	49	-29	10	6.83	1593	.0005
IPL/ang. gyrus		-35	-68	39	5.10	1377	.05						n.s.
IPS/SPL							n.s.	43	-53	48	4.59	1053	.05
precuneus		1	-59	39	4.99	2052	.001						
PCC		1	-38	21	5.06	2511	.001						
thalamus		-17	-17	15	5.59	2150		16	-15	18	4.96	1650	
cerebellum		-29	-59	-18	6.37	2700		22	-62	-15	6.58	3861	

Table 3: Activations of singing (contrasted to the control condition in which subjects did not sing). The table shows the results of the cluster analysis of statistical parametric maps ($p < 0.05$ corrected for multiple comparisons) and ROI analyses.