Specialized neural dynamics for verbal and tonal memory: fMRI evidence in congenital amusia

Philippe Albouy, Isabelle Peretz, Patrick Bermudez, Robert J Zatorre, Barbara Tillmann, Anne Caclin

Title: Specialized neural dynamics for verbal and tonal memory: fMRI evidence in congenital amusia.

Short Title: Neural dynamics for verbal and tonal memory

Authors:

Philippe Albouy[1, 2, 3, 4, 5], Isabelle Peretz[5], Patrick Bermudez[4, 5], Robert J. Zatorre[4, 5], Barbara Tillmann[1, 3]*, and Anne Caclin [2, 3]*

* Both authors contributed equally to this work

Affiliations:

[1] CNRS, UMR5292; INSERM, U1028; Lyon Neuroscience Research Center, Auditory Cognition and Psychoacoustics team, Lyon, F-69000, France

[2] INSERM, U1028; CNRS, UMR5292; Lyon Neuroscience Research Center, Brain Dynamics and Cognition Team, Lyon, F-69000, France

[3] University Lyon 1, Lyon, F-69000, France

[4] Cognitive Neuroscience Unit, Montreal Neurological Institute, McGill University

[5] International Laboratory for Brain, Music and Sound Research (BRAMS), Montreal, QC, Canada H3C 3J7

Corresponding author: Philippe Albouy, Cognitive Neuroscience Unit, Montreal Neurological Institute, McGill University, E-mail: philippe.albouy@mail.mcgill.ca

Conflict of interest statement: None declared.

Abstract

Behavioral and neuropsychological studies have suggested that tonal and verbal short-term memory are supported by specialized neural networks. To date however, neuroimaging investigations have failed to confirm this hypothesis. In this study, we investigated the hypothesis of distinct neural resources for tonal and verbal memory by comparing typical non-musician listeners to individuals with congenital amusia, who exhibit pitch memory impairments with preserved verbal memory. During fMRI, amusics and matched controls performed delayed-match-to-sample tasks with tones and words and perceptual control tasks with the same stimuli. For tonal maintenance, amusics showed decreased activity in the right auditory cortex, Inferior Frontal Gyrus (IFG) and Dorso-lateral-Prefrontal cortex (DLPFC). Moreover, they exhibited reduced right-lateralized functional connectivity between the auditory cortex and the IFG during tonal encoding and between the IFG and the DLPFC during tonal maintenance. In contrasts, amusics showed no difference compared to the controls for verbal memory, with activation in the left IFG and left fronto-temporal connectivity. Critically, we observed a group-by-material interaction in right fronto-temporal regions: while amusics recruited these regions less strongly for tonal memory than verbal memory, control participants showed the reversed pattern (tonal > verbal). By benefitting from the rare condition of amusia, our findings suggest specialized cortical systems for tonal and verbal short-term memory in the human brain.

Key words: Tone deafness, brain networks, auditory, Inferior Frontal Gyrus, memory

Introduction

Short-term memory is a cognitive ability allowing for the maintenance of information for a short period of time (seconds or minutes, see Baddeley, 2010; Cowan, 2008; D'Esposito, 2007; Logie & D'Esposito, 2007). Short-term memory is often linked to working memory (Baddeley, 2010; Cowan, 2008): while short-term memory has been used to refer to the simple temporary storage of information (maintenance), working memory refers to the maintenance and the simultaneous manipulation of information (Baddeley, 1986; Engle et al., 1999). In the present study, we focused on short-term memory by studying the maintenance of tonal and verbal information. For verbal short-term memory (i.e. words, syllables), neuroimaging studies have suggested that the cortical networks supporting rehearsal processes during maintenance are similar to those involved in speech perception and production (Buchsbaum & D'Esposito, 2008; Buchsbaum, Olsen, Koch, & Berman, 2005; Schulze & Koelsch, 2012). This conclusion derives from the observation of activations of left inferior frontal regions during short-term memory tasks requiring subvocal rehearsal as the main strategy of maintenance (Awh et al., 1996; Fiez et al., 1996; Gruber & von Cramon, 2003; Paulesu, Frith, & Frackowiak, 1993; Ravizza, Delgado, Chein, Becker, & Fiez, 2004). Furthermore, the Posterior Parietal Cortex (PPC), notably the inferior parietal lobule, and the left planum temporale, are thought to support the temporary storage of verbal information (Buchsbaum & D'Esposito, 2008; Buchsbaum et al., 2005; Hickok, Buchsbaum, Humphries, & Muftuler, 2003). In addition to the cortical networks supporting rehearsal and passive storage, other frontal regions are recruited during verbal short-term memory. Notably, the Dorso-Lateral-Prefrontal-Cortex (DLPFC) (Curtis & D'Esposito, 2003; D'Esposito, 2007; Logie & D'Esposito, 2007; Owen, 2000) has been shown to be involved in the monitoring of information stored in memory (D'Esposito, 2007; Owen, 2000; Petrides, 1991, 1994).

In contrast to verbal short-term memory, the behavioral and cerebral correlates of tonal short-term memory have been much less investigated. Using classic interference tasks, original studies have hypothesized that the temporary storage of pitch in tonal memory is supported by a specialized subsystem in the brain (Deutsch, 1970, 1975; Pechmann & Mohr, 1992). Notably, Deutsch's (1970) described that pitch memory is subject to interference from other pitch information (tones), but not from verbal information (e.g., numbers). However this hypothesis has been challenged, notably by Semal, Demany, Ueda, & Halle (1996) who have specified that pitch memory is influenced more strongly by the proximity of the pitch of the interfering sounds than by the verbal versus non-verbal nature of the interfering material. Overall, behavioral studies have suggested that tonal short-term memory might recruit cognitive mechanisms that are very similar to those involved in verbal (i.e. phonological) short-term memory (Salame & Baddeley, 1989, Schendel & Palmer, 2007; Williamson, Baddeley, & Hitch, 2010).

When investigating the neural correlates of tonal (pitch) memory, activations involving the Inferior Frontal Gyrus (IFG), the DLPFC and the insular cortex, along with the planum temporale, the Intra Parietal Sulcus (IPS), the Hippocampus, the supramarginal gyrus (SMG) and cerebellum have been reported (Albouy, Mattout, et al., 2013; Albouy, Mattout, Sanchez, Tillmann, & Caclin, 2015; Albouy, Weiss, Baillet, & Zatorre, 2017; Foster, Halpern, & Zatorre, 2013; Foster & Zatorre, 2010; Gaab, Gaser, Zaehle, Jancke, & Schlaug, 2003; Griffiths, Johnsrude, Dean, & Green, 1999; Holcomb et al., 1998; Kumar et al., 2016; Zatorre, Evans, & Meyer, 1994). While the brain regions involved in pitch memory are highly comparable to brain regions recruited for the maintenance of verbal information, several of these findings reveal more strongly right-lateralized activations and thus suggest a potential specialization of each hemisphere for different materials (Caclin & Tillmann, 2018; Peretz & Zatorre, 2005; Samson & Zatorre, 1992).

However, when directly comparing the cortical networks involved in auditory short-term memory for verbal and tonal materials either in non-musicians or musicians, hemispheric specialization failed to be confirmed (Hickok et al., 2003; Koelsch et al., 2009; Schulze & Koelsch, 2012; Schulze, Zysset, Mueller, Friederici, & Koelsch, 2011). Thus, it has been hypothesized that while verbal and tonal information recruit similar subparts of more general working memory and short-term memory brain networks, these two types of information are not necessarily *processed* in the same way in terms of network dynamics (Caclin & Tillmann, 2018; Hirel et al., 2017; Peretz & Zatorre, 2005).

A way to explore this hypothesis of distinct neural resources for the two materials (verbal, tonal) is to study individuals with congenital amusia (hereafter amusia), a lifelong disorder of music processing that cannot be explained by hearing loss, brain damage, cognitive deficiencies, and that has not been linked to speech impairments (Ayotte, Peretz, & Hyde, 2002; Peretz, 2016; Stewart, 2011; Stewart, von Kriegstein, Warren, & Griffiths, 2006; Tillmann, Albouy, & Caclin, 2015; Tillmann, Leveque, Fornoni, Albouy, & Caclin, 2016). Amusics' deficits are most pronounced along the pitch dimension, and have been traced down to impairments in pitch memory (Albouy, Mattout, et al., 2013; Albouy, Schulze, Caclin, & Tillmann, 2013; Gosselin, Jolicoeur, & Peretz, 2009; Tillmann et al., 2015; Tillmann et al., 2016; Williamson, McDonald, Deutsch, Griffiths, & Stewart, 2010; Williamson & Stewart, 2010). In contrast, amusics show normal short-term memory abilities for verbal material (Tillmann, Schulze, & Foxton, 2009; Williamson & Stewart, 2010), and more generally intact speech processing, except along the pitch dimension (Tillmann et al., 2015). The pitch memory deficit is associated with delayed magnetoencephalographic responses in bilateral IFG and Superior Temporal Gyrus (STG) during the encoding of melodies and right-lateralized functional anomalies in the DLPFC and PPC during the maintenance of the melodic information in short-term memory (Albouy, Mattout, et al., 2013). These results

highlight deficits in the pitch perception and memory network described in typical individuals (Albouy, Baillet, & Zatorre, 2018; Albouy et al., 2017; Kumar et al., 2016; Zatorre et al., 1994) and are in agreement with functional and anatomical anomalies observed in the amusic brain (see Peretz, 2016 for review). Anatomical abnormalities have been reported in the right IFG, amusics' brains showing decreased white matter concentration associated with increased grey matter concentration in this region (Hyde, Zatorre, Griffiths, Lerch, & Peretz, 2006, Albouy Mattout et al, 2013), and in the right STG (Hyde et al., 2007, see also Mandell, Schulze, & Schlaug, 2007 for abnormalities observed in the left hemisphere). The hypothesis of an abnormal right fronto-temporal pathway in the amusic brain has received support by the observation of reduced fibre connectivity in the right arcuate fasciculus (Loui, Alsop, & Schlaug, 2009). To relate anatomical anomalies to behavioural expressions, functional investigations have reported decreased right fronto-temporal connectivity between the auditory cortex and right IFG observed during pitch perception (Hyde, Zatorre, & Peretz, 2011), pitch memory (Albouy et al., 2015) and also bilaterally during resting state (Leveque et al., 2016).

Overall, behavioral and neuroimaging studies on congenital amusia suggest that an altered recruitment of right fronto-temporal regions is linked to amusics' pitch memory deficits. In contrast, amusics show preserved memory abilities for verbal material thus suggesting a neural separation of the core regions involved in tonal and verbal short-term memory. However, this potential neural separation in the amusic brain has not been investigated with neuroimaging to date.

To directly test the hypothesis of distinct neural resources for verbal and tonal memory, we used functional Magnetic Resonance Imaging (fMRI) while amusics and matched controls performed a memory task and a perception task for either verbal or tonal

materials. By contrasting memory and perception tasks, we aimed to identify the brain networks specifically related to memory maintenance in each group.

Finally, in addition to testing the hypothesis that tonal and verbal memory maintenance rely on partly distinct brain mechanisms, we investigated amusics' pitch short-term memory deficits during the encoding of tonal information, a processing step for which abnormal brain functioning has been reported (see above). The experiment was thus divided into six runs: 4 runs with tonal material (2 runs for tonal encoding, 2 runs for tonal maintenance) and 2 runs with verbal material (2 runs for verbal maintenance).

Based on previous studies, we expected to observe impaired memory performance in amusic participants relative to controls for the tone sequences, but not for the verbal sequences (see Tillmann et al. 2009). Furthermore, we predicted abnormal fronto-temporal BOLD activation and connectivity in amusics as compared to controls during tonal encoding and maintenance (Albouy, Mattout, et al., 2013; Albouy et al., 2015; Hyde et al., 2011; Leveque et al., 2016), and typical left fronto-temporal activity during verbal memory in both amusics and controls.

When contrasting verbal maintenance and tonal maintenance in the amusic group, we expected to observe an enhanced recruitment of the maintenance network for verbal material, as amusics show altered maintenance of pitch information, but preserved maintenance of verbal information. This difference should be observed specifically in the right hemisphere, where decreased activity during pitch maintenance has been reported in congenital amusia (Albouy et al., 2013). Finally, if distinct networks are recruited for tonal and verbal memory (with a right lateralization for tonal information and left lateralization for verbal information, see Peretz & Zatorre (2005)), the interaction between group and material (control [tonal > verbal]: amusics [verbal > tonal]) should also highlight this difference in the right hemisphere.

Material and Methods

The data were acquired in two different fMRI centers, in Lyon (France) and Montreal (Canada), using identical 3T Philips Achieva scanners, including the same update (R3.2.2), 32-channel head coil, and imaging parameters. Similar systems were used for stimulus presentation (see 2.6.). All participants were native French speakers.

Participants

Eighteen amusic adults and 18 non-musician controls matched for gender, age, handedness, years of education, and years of musical instruction, participated in the study (see details in supplementary Table S1). The amusic group was composed of 13 participants from Lyon and 5 from Montreal. The control group was composed of 14 participants from Lyon and 4 from Montreal. Details about the groups are presented in supplementary material. All participants had right-handed laterality and reported no history of neurological or psychiatric disease. They gave their written informed consent and received a monetary compensation for their participation. All participants were tested with standard audiometry and none of them had moderate (35 dB) or severe (more than 40 dB) peripheral hearing loss at the frequencies of interest (between 250Hz to 1000Hz). Note however that one amusic participant showed a mild hearing loss at 250 Hz in the right ear (threshold at 30dB), and one control participant showed a mild hearing loss in the left ear at 1000Hz (threshold at 30dB). All participants had been thoroughly evaluated in previous testing sessions with the Montreal Battery of Evaluation of Amusia (MBEA, see supplementary Table S1, Peretz et al. 2003). Participants were considered amusic when they scored below 23 across the six tasks of the battery (maximum score = 30), the cut-off being two standard deviations below the average of the normal population. To evaluate Pitch Discrimination Thresholds (PDT), all participants were tested with a two-alternative forced-choice task using a two-down/one-up adaptive staircase procedure (see Tillmann et al. 2009 for details). The average PDT of the amusic group

(ranging from 0.13 to 2.41 semitones) was significantly higher (worse) than that of the control group (ranging from 0.05 to 0.67 semitones, see supplementary Table S1). In agreement with previous findings (Foxton, Dean, Gee, Peretz, & Griffiths, 2004; Tillmann et al., 2016; Tillmann et al., 2009; Whiteford & Oxenham, 2017), we observed a partial overlap in pitch discrimination thresholds between amusic and control groups.

Stimuli

During fMRI acquisition, participants performed four tasks: a memory task and a perception task for piano tones, a memory task and a perception task for monosyllabic words (see Fig. 1A). As mentioned above, for the tonal tasks both encoding and maintenance were investigated, whereas for the verbal task only maintenance was investigated (Fig. 1C), so there were two times more trials for the tonal tasks. For all tasks, two sequences (of words or tones) were presented sequentially and separated by a silent delay. In the memory task, participants were required to indicate whether the two sequences were the same or different. In the perception task, they were required to ignore the first sequence and indicate whether the last two items of the second sequence were the same or different. The perception task was designed as a control condition: participants listened passively to the same stimuli (i.e., the first sequence) as the one used in the memory task, but without actively encoding the information in memory. All tasks involved two three-sound (words or tones) sequences (S1, S2), separated by a silent maintenance period of 9s. For both tonal and verbal materials, each sound had a duration of 250 ms, and the three sounds were presented successively with an inter-stimulus-interval of 0 ms.

For the tonal material, 120 different three-tone melodies (that were used as S1 for the 120 tonal trials, 60 for the memory task, 60 for the perception task, see below) were created using eight piano tones differing in pitch height (Cubase software, Steinberg), all belonging to

the key of C Major (C3, D3, E3, F3, G3, A3, B3, and C4, material from Albouy, Mattout et al., (2013)). For the verbal material, 60 different sequences were created using six monosyllabic French words: toux (/tu/ - cough), loup (/lu/ - wolf), boue (/bu/ - dirt), mou (/mu/ - soft), goût (/gu/ - taste) and pou (/pu/ - bug). The words were spoken by a female voice and the recordings were processed with STRAIGHT (Kawahara & Irino, 2004) to reach a fixed pitch of 230Hz for each of them (within the range of the piano tones used in the tonal tasks). The sounds were then equalized in loudness using MATLAB software (material adapted from Tillmann et al. 2009). The words were selected from a pool of recorded words judged as intelligible by eight native French speakers. For verbal and tonal material, half S1 sequences contained items repetition (words or tones) in the second and third position of the sequence and the other half did not contain item repetition within the sequence (Fig. 1A).

Memory Tasks

There were 60 memory trials (S1, silence, S2) for tones and 30 memory trials (S1, silence, S2) for words, each set being equally composed of 50% *same* and 50% *different* trials. For different trials, one item of the S2 sequence was different from the S1 sequence (in positions 1 to 3, equally distributed across trials). For melodies, this new item created a contour-violation in the melody. The pitch interval size between the original tone in S1 and the changed tone in S2 was above the PDT of all participants and controlled so that there were 50% of the trials with a medium interval size (of 1.5, 2 and 2.5 tones in equal proportion) and 50% of trials with a large interval size (of 3, 3.5 and 4 tones). For verbal sequences, the changed word was selected from the remaining words that were not presented in the S1 sequence.

Perception Tasks

The perception task consisted of 60 trials (S1, silence, S2) for tones and 30 trials (S1, silence, S2) for words (see Fig. 1A). Trials were divided into *same* and *different*. Importantly, S1

sequences in perception trials were not strictly identical to S1 sequence in memory trials, to avoid exact stimulus repetition, but were similar in terms of melodic contour for the tonal material. Moreover, the perception and memory trials followed the same constraints in terms of trials characteristics (amount of same and different trials, position of the new element in the different trials, pitch interval of the changed tone, etc.).

Procedure

Amusic and control participants performed the four tasks during fMRI recording. Presentation software (Neurobehavioral systems, Albany, CA, USA) was used to run the experiment and to record button presses. Stimuli were presented via MRI-compatible insert earphones (*NordicNeuroLab*, in Lyon and *Etymotic Research* in Montreal). The level of sound presentation was set to 70 dB SPL for all participants.

The experiment was divided into six runs of about 9 minutes each: 4 runs with tonal material (2 runs for tonal encoding, 2 runs for tonal maintenance) and 2 runs with verbal material (verbal maintenance). Within a run, memory and perception tasks were presented in blocks of 15 trials each and the task order was counterbalanced across runs and participants. At the beginning of each run, 5 trials of silence served as baseline. Task instructions were presented visually at the beginning and at the middle of each run. During fMRI acquisition, participants were asked to keep their eyes closed. When the task changed, participants heard a salient tone burst, looked at the visual instruction on the screen, and closed their eyes again. The runs were separated by 2-3 minutes of break. Participants were informed about the material (tones or words) and the order of the to-be-performed tasks before each run.

For each trial within a run, participants indicated their answers by pressing one of two keys of a response device with their right hand after the end of S2. They had 2 seconds to respond before the next trial, which occurred between to 2.5 s and 3.0 s after the end of S2. In

each task, trials were presented in a pseudo-randomized order with the constraint that the same trial type (same or different) could not be repeated more than three times in a row. Before entering the scanner, participants performed 15 practice trials for each task (with simulated scanner noise) with response feedback. No feedback was given during the main experiment.

fMRI design and acquisition parameters

At the beginning of the MRI session, a high-resolution 3D anatomical MPRAGE T1-weighted image was acquired for each participant using a gradient-echo sequence (160 sagittal slices; time to repetition (TR), 2800 ms; time to echo (TE) 3.8 ms; flip angle (FA), 8°; matrix size, 240x240; field of view (FOV) 240x240mm^2; voxel size, 1 x 1 x 1 mm^3).

A gradient-echo EPI pulse sequence was used to measure whole-brain blood oxygenation level-dependent (BOLD) signal (47 axial slices acquired in ascending sequential order; TR, 14000 ms; volume acquisition, TA = 3000 ms; TE, 30 ms; FA, 90°; 3 mm slice thickness; no gap; matrix size, 80x80 FOV 240 x 240mm^2; voxel size, 3 x 3 x 3 mm^3). The long TR (14 sec including 3 sec of image acquisition, TA) is related to the sparse-sampling paradigm (Fig. 1C) that was used to maximize task-related BOLD response and minimize auditory masking due to MRI scanning noise (Belin, Zatorre, Hoge, Evans, & Pike, 1999). Auditory events were synchronized with fMRI image volume acquisitions at a rate of one image per trial. Within different blocks, we aimed to capture the hemodynamic response associated with two different processes. First, the activity related to the maintenance of the tonal and verbal stimuli was measured with fMRI volumes acquired 5500 to 6000 ms after the end of S1 (Fig. 1C, upper panel), thereby decreasing the likelihood of capturing the activity related to the encoding of the S1 stimulus. In two additional runs, we measured the activity related to the encoding of the tonal stimuli (Fig. 1C lower panel, with fMRI volumes acquired 3500 to 4000 ms after the end of S1, i.e., at the expected peak of the hemodynamic response

for auditory processing of S1). The encoding scans were performed only for the tonal material (2 runs) to investigate whether amusics' altered responses observed during the encoding of melodies with MEG (Albouy, Mattout, et al., 2013) could be observed with another imaging method. Note that the maintenance scans were performed for both verbal and tonal materials (2 runs each).

Preprocessing

All image preprocessing was performed using SPM12 (Wellcome Trust Centre for Neuroimaging, http://www.fil.ion.ucl.ac.uk/spm/, London, UK). Before preprocessing, all images were checked for artifacts and automatically aligned so that the origin of the coordinate system was located at the anterior commissure. Preprocessing included the realignment of functional images and the co-registration of functional and anatomical data. We then performed a spatial normalization (voxel size, 3x3x3) of the T1 and the EPI images to the Montreal Neurological Institute templates provided with SPM12 (MNI T1 template and EPI template respectively). Finally, functional images were spatially smoothed (Gaussian kernel, 5 mm full-width at half-maximum (FWHM)).

fMRI analyses

Multicenter studies can entrain site-dependent effects in fMRI sensitivity, notably regarding activation effect sizes. Friedman & Glover (2006) have suggested that these confounding effects are mainly linked to different field strength, hardware and software used in different centers. In the present study, we used similar hardware, software, same update versions, fMRI sequences, and head coil in the two MRI centers in order to reduce the risk of scanner site effect. Moreover, note that in order to control for confounding site effects, scanner site (Lyon, Montreal) was modeled as a covariate of non-interest in all group-level fMRI analyses presented below. Individual contrast maps were first calculated for each

participant. A hemodynamic response function (HRF) was chosen to model the BOLD response such that it accounted for the long TR of 14 seconds (micro time resolution of 80 ms; micro time onset 1; high-pass filter 360-s). At the first level, for each participant, changes in brain regional responses were estimated by a general linear model (Friston et al., 1995) and the following contrasts were performed for each material: 1) memory vs. silence (the scans acquired at the beginning of each run, without any auditory stimulation), 2) perception vs. silence, and 3) memory vs. perception.

Contrasts were computed separately for encoding and maintenance scans (for tonal material). We then analyzed within and between-group effects at the second level. Statistical inferences were performed at a threshold of $p < 0.05$ after False Discovery Rate correction for multiple comparisons.

Functional connectivity

Functional connectivity analysis was performed using the CONN-fMRI toolbox for SPM (http://www.nitrc.org/projects/conn). Temporal correlations were computed between the BOLD signals from seed regions of interest (ROIs) to all other voxels in the brain. A general linear model was fitted to analyze BOLD activity of each participant for each condition (without HRF convolution). Data were band-pass filtered (0.008–0.09 Hz) and nuisance covariates were included to control for fluctuations in BOLD signal resulting from cerebrospinal fluid, white matter, and their derivatives. Eight seed ROIs from the FSL Harvard-Oxford and AAL atlases were selected based on previous studies showing differences between amusics and controls in terms of effective or functional connectivity during tonal perception and memory or at rest (Albouy, Mattout, et al., 2013; Albouy et al., 2015; Hyde et al., 2011; Leveque et al., 2016). Atlas-based seed definition was chosen so that the seed regions encompass the different coordinates of the studies reported above. These regions were bilateral Heschl's Gyri (right, x = 46, y = -17, z = 10; left, x = -42, y = -19, z =

10), bilateral IFG (opercular part, right, x = 50, y = 15, z = 21; left, x = -48, y = 13, z = 19), bilateral anterior (right, x = 62, y = -2, z = -2; left, x = -54, y = -10, z = -10) and posterior STG (right, x = 62, y = -24, z = 2; left, x = -58, y = -24, z = 2). Mean activation in these regions was regressed on a voxel-by-voxel basis to determine where activity significantly co-varied with the activity in that seed. Statistical analyses of correlations between the seeds and cortical areas within the memory network were performed in two steps: first, by performing the memory vs. perception contrasts of the correlation values at the first level and, second, by comparing groups at the second level analysis for each material. Statistical significance was established with a voxel-level threshold of $p<0.001$ with a cluster-level correction at q-FDR $p < .05$.

Results

Short-term memory deficit for tones, but not for words in congenital amusia

Task performance was evaluated using *d'* (signal detection theory[1]). Behavioral data were analyzed with a 2x2x2 ANOVA with group as between-participant factor, and material (tones, words) and task (memory, perception) as within-participant factors (see Fig. 1B). All assumption were met for the ANOVA (Shapiro-Wilk for normality, all ps > .08; and Levene's test for homogeneity of variance, all ps > .16). The main effect of group was significant ($F(1,34) = 11.91$, $p = .001$), with amusics showing decreased performance in comparison to controls. The main effect of material ($F(1,34) = 3.83$, $p = .05$) was significant as well as the interaction between material and group ($F(1,34) = 12.16$; $p = .001$). Moreover, the material-by-group-by-task interaction ($F(1,34) = 7.25$; $p = .01$) reached significance. Post-hoc tests (Tukey corrected) revealed that amusics' performance was decreased in comparison to controls for the tonal tasks (within and between tasks, all ps < .04), but not for verbal tasks

[1] when necessary for *d'* estimation, 0 was replaced by 0.01 for the number of false alarms, and 1 by 0.99 for the maximum number of hits (Macmillan and Creelman (1991))

(memory and perception, all ps > .98). For controls, performance was better for tonal material than for verbal material for the memory task ($p < .001$), but not for the perception task ($p = .49$). For amusics, performance was not significantly different for the verbal material and the tonal material in both tasks (all ps > .86).

Finally, amusics' performance in the tonal memory task was positively correlated (Pearson's) with the MBEA ($r(16) = .49$ $p = .03$), but not with the PDT ($r(16) = -.23$ $p = .37$). Additionally, amusics' performance in the tonal perception task was negatively correlated with the PDT ($r(16) = -.51$ $p = .03$), but not with the MBEA ($r(16) = .40$ $p = .09$). None of these correlations were significant in controls (all ps >.16 see supplementary Fig. S1).

Insert Figure 1 around here

fMRI

Five main sets of fMRI analyses were performed on correct trials only to investigate brain activity related to correctly performed encoding and maintenance of auditory information[2]. The first set aimed to define memory-related neural networks by examining brain activations in all participants when maintaining or encoding auditory material in memory as compared to silence trials (memory vs. silence; maintenance scans for verbal and tonal materials, encoding scans for tonal material). To confirm that the perception task did not recruit memory networks, we also computed the perception vs. silence contrasts (second set). To investigate whether these networks are differently (or similarly) recruited in amusic and control groups, we compared the groups for each material (brain activations and functional connectivity metrics) during the maintenance of verbal information (second set), and of tonal information (third set).

[2] Correct trials: Tonal task: controls 90.79 % ± 5.45 (mean ± SD), amusics 80.32 % ± 10.17 ; Verbal tasks controls 83.85 % ± 7.67, amusics 82.88 % ± 7.88

In the fourth set of analyses, we investigated potential differences between the two types of materials (verbal and tonal) in each group as well as the interaction between group and task. Finally, in addition to testing the hypothesis that tonal and verbal memory maintenance rely on partly distinct brain mechanisms, the fifth set of analysis intended to further characterize the cerebral underpinnings of amusics' pitch short-term memory deficits by investigating brain activations and connectivity metrics during tonal encoding."

Distributed networks supporting encoding and maintenance of auditory information

Fig. 1C shows the brain regions where activity was increased for memory tasks and perception tasks as compared to baseline (silence) during maintenance (tonal and verbal memory, left panel) and encoding (tonal memory, right panel) for all participants. While brain regions related to encoding in memory of tonal material (see supplementary Table S2) included mainly auditory regions, brain regions related to the maintenance of tonal and verbal materials in memory included bilateral superior- and inferior-frontal regions, in addition to primary and secondary auditory cortices. These analyses showed that participants recruited a more distributed network during maintenance than during the encoding of auditory information for memory trials (see also Supplementary Fig. S3 for a direct comparison between encoding and maintenance in each participant group separately). Interestingly, the contrast between perception trials and silence did not reveal any significant cluster for the maintenance scans. This suggests that the perception task is a proper control condition, as it did not involve memory-related brain networks. As expected, for encoding scans, bilateral auditory regions were recruited in the perception task, reflecting automatic sound processing during passive listening. Here below, we investigated whether memory-related BOLD activity for the memory vs. perception trials differs as a function of the group (amusics, controls) and as a function of the type of auditory material (tones, words).

Altered brain responses for tones, but not for words in congenital amusia

For verbal maintenance, all participants were first pooled together in the second level analysis (Fig. 2A) to show that the maintenance of verbal information in memory is supported by activity in the opercular part of the left IFG (x= -40; y= 16; z= 22; t= 4.75; $p < .05$ FDR-corrected; k= 126). Moreover, parameter estimates extracted from this region were positively correlated with participants' behavioral performance in the memory task for words $r(34) = .49$, $p < .05$. Interestingly, group comparisons did not show any significant cluster (see also Supplementary Fig. S2A). Finally, functional connectivity analysis showed that participants exhibited increased left fronto-temporal connectivity during the maintenance of the verbal sequences in memory trials as compared to perception trials (between the left anterior STG and the left IFG, x= -48; y = 18; z = 11; t = 4.63 $p < .05$ FDR-corrected, k =106 voxels, see supplementary Table S2, Fig. 2A, lower panel).

Insert Figure 2 around here

To investigate the cortical networks related to the maintenance of pitch information in memory, we first pooled all participants together in a second level analysis (memory vs. perception contrast). This analysis did not reveal any significant cluster. One hypothesis to explain this absence of effect would be that the two groups recruited different networks during memory maintenance, the indices of which would have been obscured by pooling the participants across groups (see also Supplementary Fig. S2). To test this hypothesis, we performed group comparisons at the whole brain level.

Group comparisons for the memory vs. perception contrast during tonal maintenance revealed that as compared to controls, amusics showed decreased BOLD activity in right superior frontal regions (including the right DLPFC and IFG), right temporal cortices (Fig. 2B Left panel), and left IFG (see supplementary Table S2). Moreover, activity in the right IFG

was positively correlated with behavioral performance in the memory task for tones in controls $r(16) = .80$, $p < .0001$, but not in amusics $r(16) = .25$, $p > .05$. Moreover, note that fMRI activity during tonal maintenance was not correlated to participants' pitch discrimination abilities (PDT).

Interestingly, the reverse contrast (amusics vs. controls) revealed that amusic participants had greater activity in auditory regions (see supplementary Table S2) and were showing a negative correlation between BOLD activation in the left STG and behavioral performance in the tonal task $r(16) = -.78$, $p < .001$. Amusics are thus recruiting for tonal maintenance a cortical network involving principally sensory regions. This result was confirmed by a supplementary analysis comparing encoding and maintenance scans in each group (for the memory vs. perception contrast): while control participants recruited different networks for the different processing stages (see Supplementary Fig. 2), these comparisons were not significant in amusic participants.

Finally, functional connectivity analysis showed that, as compared to controls, amusics exhibited decreased connectivity between the right IFG and the right DLPFC during the maintenance of tonal sequences in memory ($x= 24$; $y = 46$; $z = 32$; $t = 4.78$ $p < .05$ FDR-corrected, $k =30$ voxels, Fig. 3B).

Distinct networks for tonal and verbal maintenance

In line with previous reports in non-musician participants (Koelsch et al., 2009; Schulze & Koelsch, 2012; Schulze et al., 2011), the contrast between materials (tonal, verbal) in each group did not show any significant cluster using FDR correction. To highlight the potential differences between materials, we investigated the interaction between group and materials. Interestingly, the Group by Material interaction (Controls: Tones > Words and Amusics: Words > Tones, Fig. 3, $p < .05$ FDR-corrected) was significant in the right auditory cortex ($x= 48$; $y = -14$; $z = 8$; $t = 4.40$, $k =160$ voxels) as well as in the right DLPFC ($x= 42$; y

= 18; z = 32; t = 3.90, k =104 voxels) and the right IFG (x= 56; y = 18; z = 22; t = 4.01, k =31 voxels).

To further analyze this effect, parameter estimates were extracted for these regions and the differences between materials (tonal minus verbal) were analyzed with a 2x3 ANOVA with region (right auditory cortex, right IFG, right DLPFC) as within-participant factor, and group (amusics, controls) as between-participant factor.

The main effect of region, $F(2,68) = 2.98$, $p = .06$, as well as the group by region interaction $F(2,68) = 2.88$, $p = .06$ were not significant. Finally, as expected, the main effect of group was significant $F(1,34) = 17.60$, $p < .001$: while amusics showed greater BOLD activation in these regions for verbal memory as compared to tonal memory (all ps < .02), controls showed the effect in the opposite direction (tonal > verbal all ps < .02, see Fig. 3).

Insert Figure 3 around here

Altered maintenance of tonal information is associated to altered encoding in congenital amusia

In the final set of analyses, we investigated the cortical networks related to the encoding of pitch information in memory (memory vs. perception). First, all participants were pooled together in the second level analysis (Fig. 4A), which revealed that encoding of pitch information in memory (as compared to perception) recruited right-lateralized regions including the IFG, the Rolandic operculum, the Superior Temporal Sulcus, the Hippocampus, and the left superior frontal, pre-central gyrus and STS (see Supplementary Table S2 and Supplementary Fig. S2). Group comparison did not show any significant cluster.

Finally, functional connectivity analysis revealed that amusics showed decreased fronto-temporal connectivity (between the right anterior STG and right Rolandic operculum/IFG) during the encoding of the tonal sequences in memory (Memory vs.

perception contrast, x= 44; y = 0; z = 6; t = 5.03 p = .019 q-FDR-corrected, k =52 voxels, Fig. 4B).

Insert Figure 4 around here

Discussion

In the present study, we investigated the cortical networks related to tonal and verbal short-term memory. Using fMRI, we studied amusics' and controls' brain responses associated with 1) the maintenance of verbal and tonal materials in short-term memory and 2) the encoding of tonal information. As expected, behavioral results showed that amusic participants exhibited impaired performance as compared to controls for tonal short-term memory, but not for verbal information.

During tonal maintenance (memory as compared to perception), controls recruited auditory, as well as inferior and superior-frontal regions (including IFG and DLPFC) and showed increased connectivity between the right IFG and the right DLPFC. Amusics, however, recruited a cortical network that was similar to the one involved during the encoding of tonal information, encompassing mainly auditory regions. These results confirm the differential role of auditory regions (sensory regions) and the DLPFC (associative, multimodal region) in low-level encoding and in memory representation (requiring higher levels of processing), respectively (D'Esposito, 2007; Logie & D'Esposito, 2007) and suggest that altered recruitment of higher order cortical areas underpins amusics' deficits of maintenance of tonal information.

Interestingly, in contrast, amusics showed brain activations similar to controls' activations for the maintenance of verbal material in short-term memory, suggesting at least some distinct cortical networks for tonal and verbal memory. This hypothesis of distinct resources was confirmed by a group by material interaction revealing that right fronto-temporal regions were more active in amusics for verbal than for tonal memory, while the reversed pattern was observed in controls. Observing normal recruitment of high-level regions for verbal material in the amusic brain, as well as decreased representation for tonal material, suggests that tonal and verbal memory are processed with different neural dynamics in the

amusic brain. Moreover, these results suggest that amusics' deficits in recruiting high-level regions during tonal memory may be related to their deficit in low-level encoding of tones. fMRI analyses of tonal encoding were in line with this hypothesis, with amusics showing decreased fronto-temporal connectivity as compared to controls when they were actively encoding tones in memory (memory vs. perception).

Distributed networks supporting encoding and maintenance of auditory information

fMRI results in all participants during the encoding of the tonal memory task (as compared to silence, Fig. 1C) revealed a classic pattern of activity in bilateral auditory and right inferior frontal cortices (see supplementary Table S2 and supplementary Fig. S2). This activation pattern is in line with recent studies showing the role of these regions, together with functional and effective connectivity between them, in online maintenance and integration of sequential auditory events (Albouy et al., 2018; Albouy, Mattout, et al., 2013; Albouy et al., 2015; Albouy et al., 2017; Foster & Zatorre, 2010; Kumar et al., 2016; Zatorre et al., 1994). In addition to the auditory cortices and inferior frontal regions, activity emerged in bilateral frontal regions including the DLPFC and IFG during maintenance of tonal and verbal materials. This finding accords with previous data showing a strong implication of these regions in the maintenance of information in short-term memory (Logie & D'Esposito 2007) and more specifically for pitch memory processing (Peretz & Zatorre 2005; Zatorre et al. 1994); Finally, it is worth noting that while perception trials recruit bilateral auditory cortices during encoding, those trials did not show any significant cluster in comparison to silence for the maintenance. This confirms that participants were not recruiting memory networks for performing the perception task, which thus can be considered as an appropriate control condition. This assumption also finds support on the fact that amusics' PDT were negatively correlated with their performance in the tonal perception task, but not with the tonal memory task, thus confirming that the perception task requires mainly pitch discrimination and not

memory processes. In order to investigate if specific memory networks can be observed for different auditory materials, we investigated the memory vs. perception contrast for verbal and tonal memory between groups.

Short-term memory deficit and altered brain responses for tones, but not for words in congenital amusia

In comparison to controls, amusics' behavioral performance was unimpaired for verbal material, but was impaired for tonal material (Fig. 1B) for both memory and perception trials. Regarding memory, this observation agrees with the behavioral data reported by Tillmann et al. (2009), who suggested that the short-term memory deficit in congenital amusia might be pitch-specific and not affecting other memory domains. This is also in line with results showing deficits in tonal short-term memory, but normal memory spans implemented with verbal materials such as digits (Albouy, Schulze, et al., 2013; Williamson, McDonald, et al., 2010). Interestingly, controls exhibited better performance for tonal memory in comparison to verbal memory (probably benefiting from the contour information in tonal sequences, see Tillmann et al. 2009), while amusics not. Observing similar performance between the two groups for verbal information, but strongly decreased performance for the tonal information in amusics constitutes evidence for the pitch-related short-term memory deficit in congenital amusia, but not a general short-term memory deficit, which would affect verbal memory too.

In line with Tillmann et al. (2016), we propose that pitch discrimination deficits can be excluded as the sole cause of impaired short-term memory performance because: 1) the pitch thresholds of 12 out of 18 amusics were comparable to those of controls (see also Tillmann et al. 2009), 2) impaired tonal short-term memory was observed in amusics (in comparison to controls) for pitch changes corresponding to intervals that were larger than the PDTs of all amusics tested here, 3) neither amusics' nor controls' tonal memory performance was

correlated with their PDTs (analysis performed in each group separately, see Supplementary Fig. S1), and 4) fMRI activity during tonal maintenance was not correlated to participants' PDT (for each group).

To link brain activation to the behavioral expressions of the deficit, we contrasted amusics' and controls' BOLD responses during tonal and verbal maintenance. Group comparison for verbal maintenance did not reveal any significant clusters. Indeed, and as expected, both amusic and control groups showed greater activation in the left IFG during verbal memory as compared to verbal perception, and activity in this region was positively correlated with participants' behavioral performance (Fig. 2A). The observation of activity in several parts of the left IFG (pars triangularis and opercularis) for verbal memory is in line with a number of neuroimaging studies that posit a role for subvocal rehearsal mechanisms in verbal short-term memory in this region (Gruber & von Cramon, 2003; Paulesu et al., 1993; Ravizza et al., 2004). This result thus supports the dominant hypothesis suggesting the existence of similarities between the cortical networks for short-term memory for words on one hand and speech perception and production on the other. Moreover, functional connectivity between left anterior STG and left IFG was increased during verbal memory as compared to verbal perception in both groups, suggesting that short-term memory processing for verbal material in the amusic brain is preserved and recruits similar networks to those observed in controls. Preserved behavioral performance and cerebral activation during verbal short-term memory thus confirm that in congenital amusia, there is not a general short-term memory deficit.

In contrast to verbal memory, amusics showed altered brain responses for tonal memory. As compared to controls, they showed decreased activity in right auditory regions, right IFG and right DLPFC during tonal maintenance (memory vs. perception contrast, Fig. 2B left panel). The observation of decreased activity in the right DLPFC is in agreement with

Albouy, Mattout et al. (2013), whose analysis of gamma-band activity (measured with MEG) showed that while controls recruited the right DLPFC during the maintenance delay of a melodic contour task (peak at x = 45 y = 31 z = 25 in Albouy, Mattout et al. 2013 and x = 48 y = 34 z = 18 in the present study), amusics did not. The role of the right DLPFC in tonal maintenance is in line with a recent study showing that the modulation of this region with 35 Hz (gamma) transcranial Alternating Current Stimulation causally improves pitch memory performance in congenital amusia (Schaal, Pfeifer, Krause, & Pollok, 2015). Taking together these various lines of evidence and the increased connectivity between the right IFG and the right DLPFC in controls as compared to amusics observed in the present study, our data provide further support that the high-level representations of pitch information (e.g. melodic contour) is supported by the DLPFC in interaction with IFG and auditory cortices (Fig. 2B, Supplementary Fig. S2). This was confirmed by a positive correlation between activity in the right IFG and controls' behavioral performance in the tonal memory task.

In addition to regions in the right hemisphere, amusics showed decreased activity in the left IFG as compared to controls (Fig. 2B right upper panel). This agrees with previous studies suggesting that bilateral fronto-temporal pathways support the maintenance of pitch information (Albouy et al., 2017; Kumar et al., 2016; Zatorre et al., 1994), as well as a number of working memory studies showing that the left IFG is implicated in tasks requiring maintenance through articulatory rehearsal processes (Awh et al., 1996; Paulesu et al., 1993). The present results confirm that the role of the left IFG is not restricted to phonological working memory, but also extends to the rehearsal of pitch information (see Koelsch et al. 2009, Kumar et al. 2016).

Overall, these results suggest that while controls combined rehearsal strategies (involving sensory and inferior frontal regions) and high-level working-memory resources (recruiting superior frontal regions) to perform tonal maintenance, amusics did not.

Interestingly, this interpretation was confirmed by the contrast amusics versus controls, showing that during tonal maintenance, amusics recruit mainly auditory regions (see supplementary Table S2, Fig. 2B right upper panel), while the reversed contrast (controls versus amusics) reveal right frontal regions including the DLPFC and the IFG. Observing activity in auditory regions in the amusic brain during maintenance could be considered as a marker of maladaptive plasticity, as also suggested by the negative correlation between BOLD activity in left auditory cortex and behavioral performance for tonal memory in the amusic group.

fMRI studies have suggested a differential role of superior frontal and sensory and inferior frontal regions in short-term memory (D'Esposito, 2007; Logie & D'Esposito, 2007; Owen, 2000). It has been proposed that the sensory and inferior frontal regions play a general role in memory, notably by triggering active low-level encoding strategies (Owen, 2000). In contrast, superior frontal regions (such as the DLPFC) have been hypothesized to mediate more complex types of processing. The DLPFC could be considered as a specialized region where stimuli or events, previously encoded and maintained in other association cortical areas, can be re-coded, monitored, and manipulated (D'Esposito, 2007). Observing that, for tonal material, amusics recruit mostly brain areas that are typically more strongly involved in low-level memory processes (encoding), in addition to their reduced responses in regions requiring high-level memory mechanisms (monitoring and manipulation for the DLPFC), let us argue that amusics' short-term memory deficit is related at least partly to the transformation of tonal information into high-level memory representation (e.g. computation of pitch contour).

Distinct networks for tonal and verbal maintenance

Overall, we the results described above show comparable cerebral networks in amusics and controls for verbal material, but altered brain activity in amusics for tonal material. This provides evidence of the specificity of the short-term memory deficit for pitch in the amusic brain, and by extension suggests that these mechanisms are inherently dissociable. This hypothesis was directly addressed by the interaction between group and materials (tonal, verbal) (Fig. 3) highlighting right temporo-frontal regions. In amusics, the right fronto-temporal network including high-level (DLPFC) cortical regions were less strongly recruited during impaired tonal memory than during unimpaired verbal maintenance. This result can reflect impaired use of working memory networks for tonal memory in the amusic brain, but could also be interpreted as an over-recruitment of right fronto-temporal regions for verbal memory. This result suggests that in order to achieve normal levels of verbal retention, amusics might compensate with an additional dependence on working memory networks in the right hemisphere. This potential mechanism of compensatory plasticity constitutes supplementary evidence for the atypical function of the right fronto-temporal pathway in congenital amusia.

By contrast with amusics, controls recruit the right fronto-temporal network more for tonal than verbal memory. Based on these results, we can conclude that while controls may use similar networks for verbal and tonal maintenance, they tend to recruit more right lateralized regions for tonal memory. This observation, combined with the observed unimpaired recruitment of high-level regions in the left hemisphere for verbal material in the brains of amusics, as well as decreased recruitment of right-hemispheric structures for tonal material, suggests that tonal and verbal memory may be processed with different neural dynamics. This conclusion was confirmed by the lateralization of functional connectivity effects that were preserved in the left hemisphere for verbal memory in the amusic brain, but decreased in the right hemisphere as compared to controls for tonal memory. Overall, our

results confirmed Peretz's and Zatorre's hypothesis (2005) that while verbal and tonal information share several subparts of more general working memory and short-term memory brain networks, these two types of information are not dynamically processed in the same way.

Altered maintenance of tonal information is associated to altered encoding in congenital amusia

In the present study, in addition to testing the hypothesis that tonal and verbal memory maintenance rely on partly distinct brain mechanisms (via the comparison of tonal and verbal maintenance in amusics and matched controls, as discussed above), we also intended to further characterize the cerebral underpinnings of amusics' pitch short-term memory deficits. Using MEG, amusics' deficits have been reported to start already during pitch encoding (Albouy, Mattout, et al., 2013; Albouy et al., 2015). Here, fMRI analyses during encoding revealed that both amusics and controls recruited bilateral temporal regions as well as right IFG and right Hippocampus during tonal encoding (memory vs. perception, Fig. 4). This result suggests that when encoding melodies in memory, as compared to simple perception, amusics showed normal BOLD activity in the ventral auditory pathway (see Norman-Haignere et al. 2016 for similar conclusion regarding the auditory cortex). Activity in the hippocampus is in line with a recent study (Kumar et al., 2016) showing this region is involved in the analysis of auditory stimuli in real time during encoding, via auditory-hippocampal connections.

In contrast to these fMRI data, MEG data revealed altered responses in amusics' fronto-temporal pathway (Albouy, Mattout, et al., 2013): amusics exhibiting delayed responses (between 20 to 40 ms) in comparison to controls. While these fine temporal differences could be well estimated using the high temporal resolution imaging technique of

MEG, they were not measurable with low temporal resolution imaging methods of the present study (Norman-Haignere et al., 2016).

Nevertheless, the present fMRI data congruently demonstrate that the functional connectivity between the right IFG and the right anterior STG was decreased in congenital amusia (Fig. 4), even though the overall level of activity in the right IFG during pitch encoding was comparable between the two groups (Supplementary Fig. S2). This result is in agreement with anatomical data (Albouy, Mattout, et al., 2013; Hyde et al., 2007; Hyde et al., 2006; Loui et al., 2009) and functional data (Albouy, Mattout, et al., 2013; Hyde et al., 2011; Leveque et al., 2016) showing that abnormalities in the right IFG as well as impaired functional and effective (backward) connectivity between frontal and auditory regions are linked to amusics' pitch encoding deficits, as measured behaviourally (see Tillmann, et al., 2016; Peretz, 2016 for reviews). Furthermore, these results suggest that the right IFG plays a crucial role in supporting pitch encoding in the typical brain.

Overall the encoding scans confirmed the role of the connectivity between right temporal regions and right IFG in supporting online integration of sequential pitch events in short-term memory. The deficit in low-level encoding of tones in congenital amusia might be the 'upstream' precursor to the deficit observed later for high-level representation during tonal maintenance. This hypothesis is in line with a recent study, where increasing the time available to encode pitch information resulted in preserved tonal maintenance in congenital amusia (Albouy, Cousineau, Caclin, Tillmann, & Peretz, 2016).

Conclusion

This study provides evidence for specialized cortical dynamics for tonal and verbal short-term memory in the human brain and improves our understanding of the neural

underpinnings supporting amusics' deficits. The findings confirmed the behavioral memory deficits in congenital amusia for tonal material and showed that the brain mechanisms supporting the maintenance of verbal information seem to be preserved in this developmental disorder.

Acknowledgements: This work was supported by a grant from “Agence Nationale de la Recherche” (ANR) of the French Ministry of Research ANR-11-BSH2-001-01 to BT and AC and by a grant from the Canadian Institutes of Health Research to IP. PA was funded by a PhD fellowship of the CNRS and the ERASMUS MUNDUS Auditory Cognitive Neuroscience Network. This work was conducted in the framework of the LabEx CeLyA ("Centre Lyonnais d'Acoustique", ANR-10-LABX-0060) and of the LabEx Cortex ("Construction, Function and Cognitive Function and Rehabilitation of the Cortex", ANR-11-LABX-0042) of Université de Lyon, within the program "Investissements d'avenir" (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR). We thank Yohana Lévêque, Lesly Fornoni and Mihaela Felezeu for their contribution in the recruitment of amusic participants in Lyon and Montreal.

References

Albouy, P., Baillet, S., & Zatorre, R. J. (2018). Driving working memory with frequency-tuned noninvasive brain stimulation. *Ann N Y Acad Sci*. doi:10.1111/nyas.13664

Albouy, P., Cousineau, M., Caclin, A., Tillmann, B., & Peretz, I. (2016). Impaired encoding of rapid pitch information underlies perception and memory deficits in congenital amusia. *Sci Rep, 6*, 18861. doi:10.1038/srep18861

Albouy, P., Mattout, J., Bouet, R., Maby, E., Sanchez, G., Aguera, P. E., . . . Tillmann, B. (2013). Impaired pitch perception and memory in congenital amusia: the deficit starts in the auditory cortex. *Brain, 136*(Pt 5), 1639-1661. doi:10.1093/brain/awt082

Albouy, P., Mattout, J., Sanchez, G., Tillmann, B., & Caclin, A. (2015). Altered retrieval of melodic information in congenital amusia: insights from dynamic causal modeling of MEG data. *Front Hum Neurosci, 9*, 20. doi:10.3389/fnhum.2015.00020

Albouy, P., Schulze, K., Caclin, A., & Tillmann, B. (2013). Does tonality boost short-term memory in congenital amusia? *Brain Res, 1537*, 224-232. doi:10.1016/j.brainres.2013.09.003

Albouy, P., Weiss, A., Baillet, S., & Zatorre, R. J. (2017). Selective Entrainment of Theta Oscillations in the Dorsal Stream Causally Enhances Auditory Working Memory Performance. *Neuron, 94*(1), 193-206 e195. doi:10.1016/j.neuron.2017.03.015

Awh, E., Jonides, J., Smith, E. E., Schumacher, E. H., Koeppe, R. A., & Katz, S. (1996). Dissociation of storage and rehearsal in verbal working memory: evidence from positron emission tomography. *Psychol. Sci., 7*, 25–31.

Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusia: a group study of adults afflicted with a music-specific disorder. *Brain, 125*(Pt 2), 238-251.

Baddeley, A. (2010). Working memory. *Curr Biol, 20*(4), R136-140. doi:10.1016/j.cub.2009.12.014

Belin, P., Zatorre, R. J., Hoge, R., Evans, A. C., & Pike, B. (1999). Event-related fMRI of the auditory cortex. *Neuroimage, 10*(4), 417-429. doi:10.1006/nimg.1999.0480

S1053-8119(99)90480-9 [pii]

Buchsbaum, B. R., & D'Esposito, M. (2008). The search for the phonological store: from loop to convolution. *J Cogn Neurosci, 20*(5), 762-778. doi:10.1162/jocn.2008.20501

Buchsbaum, B. R., Olsen, R. K., Koch, P., & Berman, K. F. (2005). Human dorsal and ventral auditory streams subserve rehearsal-based and echoic processes during verbal working

memory. *Neuron, 48*(4), 687-697. doi:S0896-6273(05)00889-5 [pii] 10.1016/j.neuron.2005.09.029

Caclin, A., & Tillmann, B. (2018). Musical and verbal short-term memory: insights from neurodevelopmental and neurological disorders. *Ann N Y Acad Sci, in press*.

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Prog Brain Res, 169*, 323-338. doi:10.1016/S0079-6123(07)00020-9

Curtis, C. E., & D'Esposito, M. (2003). Persistent activity in the prefrontal cortex during working memory. *Trends Cogn Sci, 7*(9), 415-423.

D'Esposito, M. (2007). From cognitive to neural models of working memory. *Philos Trans R Soc Lond B Biol Sci, 362*(1481), 761-772. doi:10.1098/rstb.2007.2086

Deutsch, D. (1970). Tones and numbers: specificity of interference in immediate memory. *Science, 168*(3939), 1604-1605.

Deutsch, D. (1975). Auditory memory. *Can J Psychol, 29*(2), 87-105.

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. *J Neurosci, 16*(2), 808-822.

Foster, N. E., Halpern, A. R., & Zatorre, R. J. (2013). Common parietal activation in musical mental transformations across pitch and time. *Neuroimage, 75*, 27-35. doi:10.1016/j.neuroimage.2013.02.044

Foster, N. E., & Zatorre, R. J. (2010). A role for the intraparietal sulcus in transforming musical pitch information. *Cereb Cortex, 20*(6), 1350-1359. doi:10.1093/cercor/bhp199

Foxton, J. M., Dean, J. L., Gee, R., Peretz, I., & Griffiths, T. D. (2004). Characterization of deficits in pitch perception underlying 'tone deafness'. *Brain, 127*(Pt 4), 801-810. doi:10.1093/brain/awh105

awh105 [pii]

Friedman, L., & Glover, G.H. (2006) Report on a multicenter fMRI quality assurance protocol. J Magn Reson Imaging; 23, 827-39.

Friston, K., Holmes, A., Worsley, K. J., Poline, J. B., Frith, C. D., & Frackowiak, R. S. J. (1995). Statistical parametric maps in functional imaging: a general linear approach. *Hum. Brain Mapp, 2*, 189- 210.

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. doi:S1053811903002246 [pii]

Gosselin, N., Jolicoeur, P., & Peretz, I. (2009). Impaired memory for pitch in congenital amusia. *Ann N Y Acad Sci, 1169*, 270-272. doi:10.1111/j.1749-6632.2009.04762.x

Griffiths, T. D., Johnsrude, I., Dean, J. L., & Green, G. G. (1999). A common neural substrate for the analysis of pitch and duration pattern in segmented sound? *Neuroreport, 10*(18), 3825-3830.

Gruber, O., & von Cramon, D. Y. (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. doi:S1053811903000892 [pii]

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. doi:10.1162/089892903322307393

Hirel, C., Nighoghossian, N., Leveque, Y., Hannoun, S., Fornoni, L., Daligault, S., . . . Caclin, A. (2017). Verbal and musical short-term memory: Variety of auditory disorders after stroke. *Brain Cogn, 113*, 10-22. doi:10.1016/j.bandc.2017.01.003

Holcomb, H. H., Medoff, D. R., Caudill, P. J., Zhao, Z., Lahti, A. C., Dannals, R. F., & Tamminga, C. A. (1998). Cerebral blood flow relationships associated with a difficult tone recognition task in trained normal volunteers. *Cereb Cortex, 8*(6), 534-542.

Hyde, K. L., Lerch, J. P., Zatorre, R. J., Griffiths, T. D., Evans, A. C., & Peretz, I. (2007). Cortical thickness in congenital amusia: when less is better than more. *J Neurosci, 27*(47), 13028-13032. doi:10.1523/JNEUROSCI.3039-07.2007

Hyde, K. L., Zatorre, R. J., Griffiths, T. D., Lerch, J. P., & Peretz, I. (2006). Morphometry of the amusic brain: a two-site study. *Brain, 129*(Pt 10), 2562-2570. doi:awl204 [pii]

10.1093/brain/awl204

Hyde, K. L., Zatorre, R. J., & Peretz, I. (2011). Functional MRI evidence of an abnormal neural network for pitch processing in congenital amusia. *Cereb Cortex, 21*(2), 292-299. doi:10.1093/cercor/bhq094

Kawahara, H., & Irino, T. (2004). Underlying principles of a high-quality speech manipulation system STRAIGHT and its application to speech segregation. . In P. L. Divenyi (Ed.), *Speech separation by humans and machines* (pp. 167–180): Kluwer Academic.

Koelsch, S., Schulze, K., Sammler, D., Fritz, T., Muller, K., & Gruber, O. (2009). Functional architecture of verbal and tonal working memory: an FMRI study. *Hum Brain Mapp, 30*(3), 859-873. doi:10.1002/hbm.20550

Kumar, S., Joseph, S., Gander, P. E., Barascud, N., Halpern, A. R., & Griffiths, T. D. (2016). A Brain System for Auditory Working Memory. *J Neurosci, 36*(16), 4492-4505. doi:10.1523/JNEUROSCI.4341-14.2016

Leveque, Y., Fauvel, B., Groussard, M., Caclin, A., Albouy, P., Platel, H., & Tillmann, B. (2016). Altered intrinsic connectivity of the auditory cortex in congenital amusia. *J Neurophysiol, 116*(1), 88-97. doi:10.1152/jn.00663.2015

Logie, R. H., & D'Esposito, M. (2007). Working memory in the brain. *Cortex, 43*(1), 1-4.

Loui, P., Alsop, D., & Schlaug, G. (2009). Tone deafness: a new disconnection syndrome? *J Neurosci, 29*(33), 10215-10220. doi:29/33/10215 [pii]

10.1523/JNEUROSCI.1701-09.2009

Macmillan, N. A., & Creelman, C. D. (1991). Detection theory: A user's guide. New York, NY, US: Cambridge University Press.

Mandell, J., Schulze, K., & Schlaug, G. (2007). Congenital amusia: an auditory-motor feedback disorder? *Restor Neurol Neurosci, 25*(3-4), 323-334.

Norman-Haignere, S. V., Albouy, P., Caclin, A., McDermott, J. H., Kanwisher, N. G., & Tillmann, B. (2016). Pitch-Responsive Cortical Regions in Congenital Amusia. *J Neurosci, 36*(10), 2986-2994. doi:10.1523/JNEUROSCI.2705-15.2016

Owen, A. M. (2000). The role of the lateral frontal cortex in mnemonic processing: the contribution of functional neuroimaging. *Exp Brain Res, 133*(1), 33-43. doi:10.1007/s002210000398

Paulesu, E., Frith, C. D., & Frackowiak, R. S. (1993). The neural correlates of the verbal component of working memory. *Nature, 362*(6418), 342-345. doi:10.1038/362342a0

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

Peretz, I. (2016). Neurobiology of Congenital Amusia. *Trends Cogn Sci, 20*(11), 857-867. doi:10.1016/j.tics.2016.09.002

Peretz, I., & Zatorre, R. J. (2005). Brain organization for music processing. *Annu Rev Psychol, 56*, 89-114. doi:10.1146/annurev.psych.56.091103.070225

Petrides, M. (1991). Monitoring of selections of visual stimuli and the primate frontal cortex. *Proc Biol Sci, 246*(1317), 293-298. doi:10.1098/rspb.1991.0157

Petrides, M. (1994). Frontal lobes and behaviour. *Curr Opin Neurobiol, 4*(2), 207-211.

Ravizza, S. M., Delgado, M. R., Chein, J. M., Becker, J. T., & Fiez, J. A. (2004). Functional dissociations within the inferior parietal cortex in verbal working memory. *Neuroimage, 22*(2), 562-573. doi:10.1016/j.neuroimage.2004.01.039

S1053811904000849 [pii]

Salame, P., & Baddeley, A. (1989). Effects of background music on phonological short-term memory. *The Quarterly Journal of Experimental Psychology, 41A*, 107-122.

Samson, S., & Zatorre, R. J. (1992). Learning and retention of melodic and verbal information after unilateral temporal lobectomy. *Neuropsychologia, 30*(9), 815-826. doi:0028-3932(92)90085-Z [pii]

Schaal, N. K., Pfeifer, J., Krause, V., & Pollok, B. (2015). From amusic to musical?--Improving pitch memory in congenital amusia with transcranial alternating current stimulation. *Behav Brain Res, 294*, 141-148. doi:10.1016/j.bbr.2015.08.003

Schendel, Z. A., & Palmer, C. (2007). Suppression effects on musical and verbal memory. *Mem Cognit, 35*(4), 640-650.

Schulze, K., & Koelsch, S. (2012). Working memory for speech and music. *Ann N Y Acad Sci, 1252*, 229-236. doi:10.1111/j.1749-6632.2012.06447.x

Schulze, K., Zysset, S., Mueller, K., Friederici, A. D., & Koelsch, S. (2011). Neuroarchitecture of verbal and tonal working memory in nonmusicians and musicians. *Hum Brain Mapp, 32*(5), 771-783. doi:10.1002/hbm.21060

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

Stewart, L. (2011). Characterizing congenital amusia. *Q J Exp Psychol (Hove), 64*(4), 625-638. doi:932314953 [pii]

10.1080/17470218.2011.552730

Stewart, L., von Kriegstein, K., Warren, J. D., & Griffiths, T. D. (2006). Music and the brain: disorders of musical listening. *Brain, 129*(Pt 10), 2533-2553. doi:awl171 [pii]

10.1093/brain/awl171

Tillmann, B., Albouy, P., & Caclin, A. (2015). Congenital amusias. *Handb Clin Neurol, 129*, 589-605. doi:10.1016/B978-0-444-62630-1.00033-0

Tillmann, B., Leveque, Y., Fornoni, L., Albouy, P., & Caclin, A. (2016). Impaired short-term memory for pitch in congenital amusia. *Brain Res, 1640*(Pt B), 251-263. doi:10.1016/j.brainres.2015.10.035

Tillmann, B., Schulze, K., & Foxton, J. M. (2009). Congenital amusia: a short-term memory deficit for non-verbal, but not verbal sounds. *Brain Cogn, 71*(3), 259-264. doi:S0278-2626(09)00156-0 [pii]

10.1016/j.bandc.2009.08.003

Whiteford, K. L., & Oxenham, A. J. (2017). Auditory deficits in amusia extend beyond poor pitch perception. *Neuropsychologia, 99*, 213-224. doi:10.1016/j.neuropsychologia.2017.03.018

Williamson, V. J., Baddeley, A. D., & Hitch, G. J. (2010). Musicians' and nonmusicians' short-term memory for verbal and musical sequences: comparing phonological similarity and pitch proximity. *Mem Cognit, 38*(2), 163-175. doi:38/2/163 [pii]

10.3758/MC.38.2.163

Williamson, V. J., McDonald, C., Deutsch, D., Griffiths, T. D., & Stewart, L. (2010). Faster decline of pitch memory over time in congenital amusia. *Adv Cogn Psychol, 6*, 15-22. doi:10.2478/v10053-008-0073-5

Williamson, V. J., & Stewart, L. (2010). Memory for pitch in congenital amusia: beyond a fine-grained pitch discrimination problem. *Memory, 18*(6), 657-669. doi:925630205 [pii]

10.1080/09658211.2010.501339

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

Figure Captions

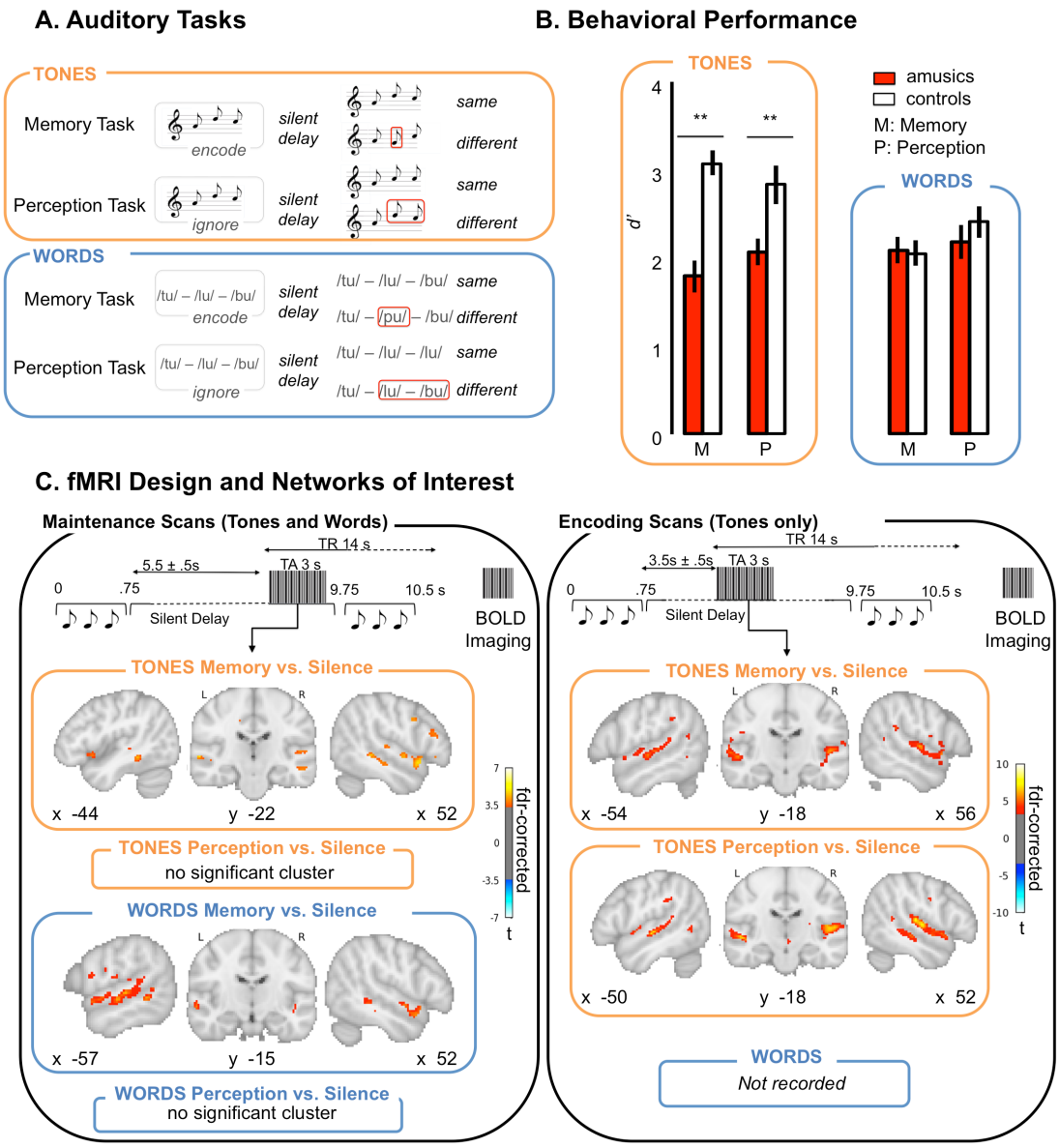
Fig. 1. A. Examples of the stimuli used in the memory and perception tasks. **B.** Performance of amusic and control groups (white, controls; red, amusics) in terms of d', presented as a function of material (orange, tonal; blue, verbal) and task (M: memory task; P: perception task). Error bars indicate s.e.m. **C**. Design for the fMRI experiment, the sparse sampling protocol, timeline of events during one trial, and brain activity for all participants. **Left panel:** For maintenance runs, the volume acquisition occurred just before the second sequence (at the end of the silent delay), the acquisition thus starting from 5500 to 6000 ms after the end of S1. Sections show brain regions where activation was increased during maintenance in memory trials (tonal top panel and verbal lower panel) as compared to baseline (silence) in all participants. FDR corrected $p < .05$. The comparison between perception trials and baseline did not show any significant cluster. These scans were performed for both tonal and verbal trials. **Right panel:** For encoding runs (two runs, tonal material only), acquisition started 3500 to 4000 ms after the end of the S1 sequence. Sections show brain regions where activity was increased during encoding in memory trials (tonal material) as compared to baseline (silence) and perception trials vs. baseline (silence) in all participants. FDR corrected $p < .05$. Results are displayed on the single subject T1 image in the MNI space provided by SPM12.

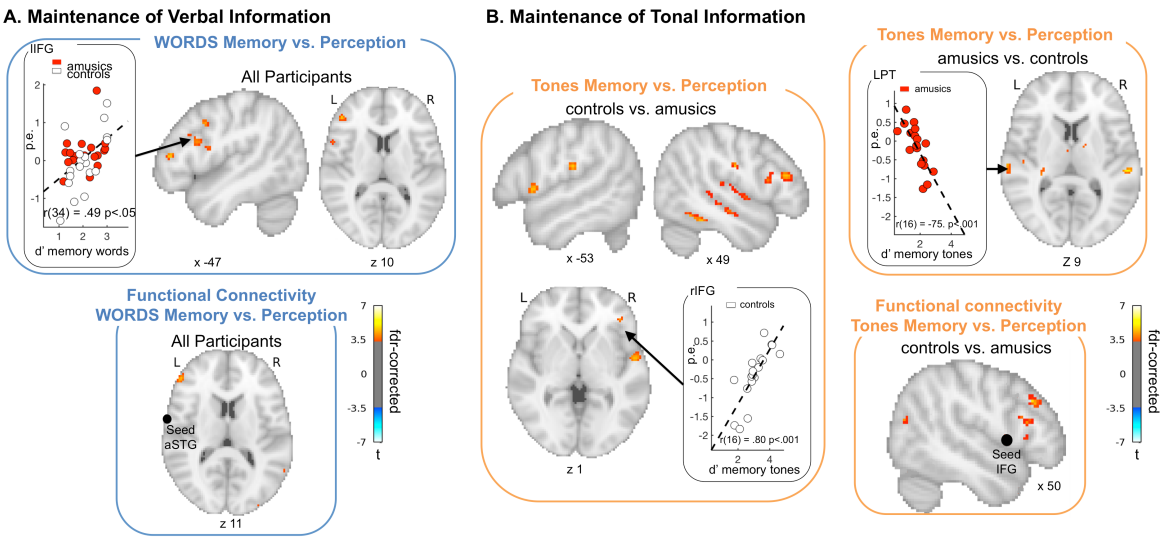
Fig. 2: Functional imaging results. **A.** Maintenance scans for verbal material (blue squares). **Top Panel** Memory vs. perception for all participants FDR-corrected $p < .05$. Scatter plot represents parameter estimates (p.e.) extracted from the left IFG for each group (red: amusics, white: controls) as a function of behavioral performance in the memory task for words; **Bottom Panel.** Seed-to-Voxel functional connectivity results for the contrast memory vs. perception (verbal), all participants. Black dot indicates the seed region in the left anterior STG. **B**. Maintenance scans for tonal material (orange squares)**. Left Panel**: controls vs. amusics (memory vs. perception) FDR-corrected $p < .05$. Scatter plot represents parameter estimates (p.e.) extracted from the right IFG for the control group (white circles) as a function of their behavioral performance in the memory task for tones. **Right top Panel**: amusics vs. controls (memory vs. perception). Scatter plot represents parameter estimates (p.e.) extracted from the left STG for the amusic group (red circles) as a function of their behavioral performance in the memory task for tones. **Right lower Panel**: Seed-to-Voxel functional connectivity results for the contrast controls vs. amusics (memory vs. perception tonal). Black dot indicates the seed region in the right IFG.

All results are displayed on the single subject T1 in the MNI space provided by SPM12. The areas of activation are detailed in supplementary Table S2.

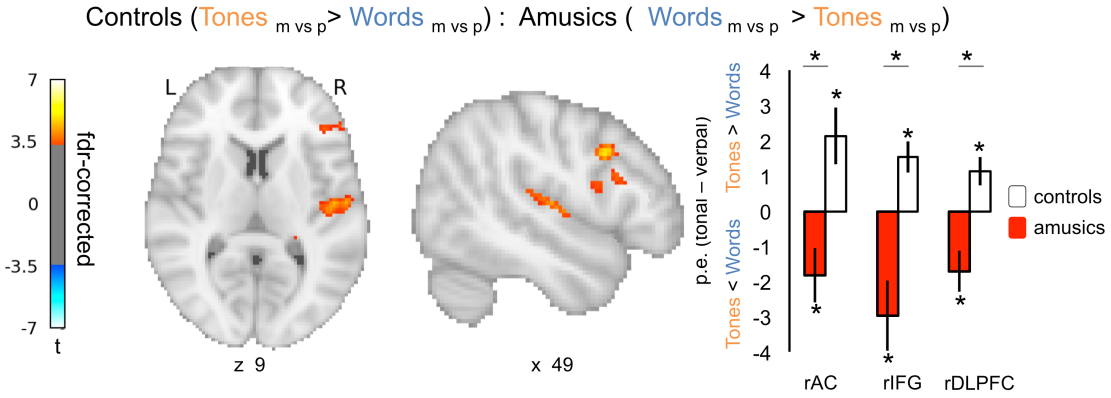
Fig. 3. Functional imaging results, maintenance scans. **A.** Group by material Interaction, $p < .05$ FDR-corrected. Memory vs. perception contrasts were performed at the first level each material for all participants. Bar plots represent parameter estimates for the difference tonal minus verbal for significant regions for each group (red, amusics; white, controls). Errors bars represent the s.e.m.

Fig. 4. Functional imaging results, encoding scans for tonal material. A. Memory vs. perception for all participants FDR-corrected $p < .05$. Results are displayed single subject T1 in the MNI space provided by SPM12. The areas of activation are detailed in supplementary Table S2. Errors bars represent the s.e.m. B. Seed to voxel functional connectivity results for the contrast controls > amusics (memory vs. perception tonal). Black dot indicates the seed region in the right anterior STG.

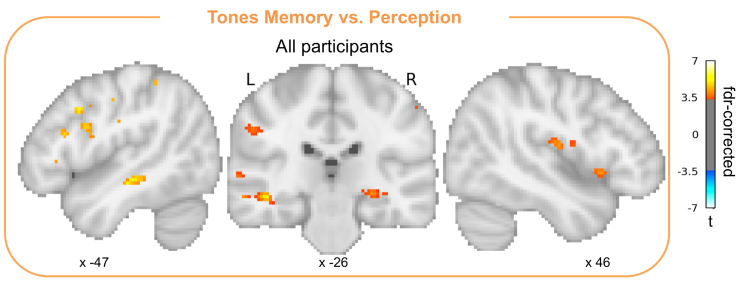

A. Auditory Tasks
TONES
Memory Task
encode
silent delay
same
different
Perception Task
ignore
silent delay
same
different
WORDS
Memory Task
/tu/ – /lu/ – /bu/
encode
silent delay
/tu/ – /lu/ – /bu/ same
/tu/ – /pu/ – /bu/ different
Perception Task
/tu/ – /lu/ – /bu/
ignore
silent delay
/tu/ – /lu/ – /lu/ same
/tu/ – /lu/ – /bu/ different
B. Behavioral Performance
TONES
d'
**
**
M
P
amusics
controls
M: Memory
P: Perception
WORDS
M
P
C. fMRI Design and Networks of Interest
Maintenance Scans (Tones and Words)
TR 14 s
5.5 ± .5s
TA 3 s
0
.75
9.75
10.5 s
Silent Delay
BOLD Imaging
TONES Memory vs. Silence
x -44
y -22
x 52
fdr-corrected
t
TONES Perception vs. Silence
no significant cluster
WORDS Memory vs. Silence
x -57
y -15
x 52
WORDS Perception vs. Silence
no significant cluster
Encoding Scans (Tones only)
TR 14 s
3.5s ± .5s
TA 3 s
0
.75
9.75
10.5 s
Silent Delay
BOLD Imaging
TONES Memory vs. Silence
x -54
y -18
x 56
fdr-corrected
t
TONES Perception vs. Silence
x -50
y -18
x 52
WORDS
Not recorded


A. Maintenance of Verbal Information
WORDS Memory vs. Perception
lIFG
amusics
controls
p.e.
r(34) = .49 p<.05
d' memory words
All Participants
L
R
x -47
z 10
Functional Connectivity
WORDS Memory vs. Perception
All Participants
L
R
Seed
aSTG
z 11
fdr-corrected
t
B. Maintenance of Tonal Information
Tones Memory vs. Perception
controls vs. amusics
x -53
x 49
L
R
rIFG
controls
r(16) = .80 p<.001
d' memory tones
z 1
Tones Memory vs. Perception
amusics vs. controls
LPT
amusics
r(16) = -75. p<.001
d' memory tones
L
R
Z 9
Functional connectivity
Tones Memory vs. Perception
controls vs. amusics
Seed
lFG
x 50
fdr-corrected
t

Maintenance: Group by Material Interaction



A. Encoding of Tonal Information



B. Functional connectivity: Encoding