The Effect of Familiarity on Neural Representations of Music and Language

Avital Sternin, Lucy M. McGarry, Adrian M. Owen, and Jessica A. Grahn

Abstract

■ We investigated how familiarity alters music and language processing in the brain. We used fMRI to measure brain responses before and after participants were familiarized with novel music and language stimuli. To manipulate the presence of language and music in the stimuli, there were four conditions: (1) whole music (music and words together), (2) instrumental music (no words), (3) a capella music (sung words, no instruments), and (4) spoken words. To manipulate participants' familiarity with the stimuli, we used novel stimuli and a familiarization paradigm designed to mimic "natural" exposure, while controlling for autobiographical memory confounds. Participants completed two fMRI scans that were separated by a stimulus training period. Behaviorally, participants learned the stimuli over the training period. However, there were no significant neural differences between the familiar and unfamiliar stimuli in either univariate or multivariate analyses. There were differences in neural activity in frontal and temporal regions based on the presence of language in the stimuli, and these differences replicated across the two scanning sessions. These results indicate that the way we engage with music is important for creating a memory of that music, and these aspects, over and above familiarity on its own, may be responsible for the robust nature of musical memory in the presence of neurodegenerative disorders such as Alzheimer's disease. ■

INTRODUCTION

Music and language abilities are closely related. At the sensory level, both music and language involve acoustic stimuli arranged in structurally meaningful ways. For example, both involve small units (music notes or words) that are combined using specific rules to create larger units (melodies/songs and sentences/stories). Cognitively, the comprehension of both music and language involves creating expectations about what comes next in a series of sounds (Patel, 2008), using learned rules (e.g., syntax) to interpret the input (Jackendoff, 2009; Jackendoff & Lerdahl, 2006), and requires the use of memory (Zatorre & Gandour, 2008; Daneman & Merikle, 1996). Although they rely on similar processes, there is evidence to suggest that both overlapping and distinct networks are involved in music and language.

Perceiving music and language activates overlapping brain networks. EEG data show that words and music are closely related in the early stages of cognitive processing (within the first 300–500 msec following the perception of the sound; Gordon, Schön, Magne, Astésano, & Besson, 2010), and fMRI studies provide evidence for anatomically similar networks. For example, Broca's area, the superior temporal sulcus, the superior temporal gyrus, the insula, and the frontal pole are known to be involved in the language network, and these areas are also active in music processing (Hymers et al., 2015; Merrill et al., 2012; Schön

et al., 2010; Fadiga, Craighero, & D'Ausilio, 2009; Koelsch et al., 2002). One cognitive ability common to both music and language is memory. In the short term, language and music unfold over time and therefore require the initial inputs to be held in mind for subsequent inputs to be understood (Peretz & Zatorre, 2005). For example, to understand the end of this sentence, you need to be able to remember what the beginning of the sentence was about (Daneman & Merikle, 1996). In music, individual or groups of notes need to be remembered to make sense of a melody. Long-term memory of music and language is responsible for our ability to sing along with a song or recite a poem from memory. Similar to a "word lexicon" that stores all of the words that we know (Mohanan, 1982), there is evidence for a "musical lexicon" (Peretz & Coltheart, 2003) that contains representations of the music that we know. Long-term memory results from mapping a perceived sound, whether it is a melody or a sentence, onto a stored representation and making a decision about whether the sound is new or not. Therefore, both music and memory rely on short-term auditory memory to make sense of the components of a sound sequence, and on long-term stored representations to judge whether an incoming sequence is novel.

Evidence for distinct music and language networks is largely driven by clinical case studies of patients with a deficit in either music or language abilities that leaves the other ability intact. For example, individuals with acquired or congenital amusia recognize spoken words and lyrics but are unable to recognize tunes and melodies

The University of Western Ontario, London, Canada

(Ayotte, Peretz, & Hyde, 2002; Piccirilli, Sciarma, & Luzzi, 2000; Griffiths, 1997; Peretz et al., 1994). The opposite deficit also exists. Some individuals with brain damage may have verbal agnosia (word deafness) and are unable to recognize spoken words but are able to recognize non-verbal sounds, including music (Takahashi et al., 1992; Yaqub, Gascon, Nosha, & Whitaker, 1988; Metz-Lutz & Dahl, 1984). Further evidence for distinct networks for music and language comes from the speech–song illusion, in which the repetition of a spoken phrase creates the perception of a song (Deutsch, Henthorn, & Lapidis, 2011). One study found that distinct areas in the fronto-temporal cortices are active when the repeated phrases are perceived as song but not when the same phrases are perceived as speech (Tierney, Dick, Deutsch, & Sereno, 2013) indicating that the neural difference is based on the perception of the phrase as language or music. Most recently, electrocorticography measured from the temporal cortex of individuals listening to a variety of sounds (e.g., birds chirping, individuals speaking, and music) found groups of neurons that responded specifically to songs and were distinct from those neurons that responded to language (Norman-Haignere et al., 2019).

There is also compelling evidence for a musical memory system that is distinct from that for language, despite the similar role that memory plays in these two abilities. Two patients with medial and lateral temporal lobe damage demonstrated severe deficits in visual and verbal memory, but intact musical memory (Esfahani-Bayerl, Finke, Kopp, Moon, & Ploner, 2019; Finke, Esfahani, & Ploner, 2012), whereas a third patient experienced the opposite deficit: intact verbal memory with a severe, music-specific agnosia (Peretz, 1996). Musical memory is also spared in some individuals with neurodegenerative disorders such as Alzheimer's disease, even in the context of deteriorating semantic memories (Slattery et al., 2019; Jacobsen, Fritz, Stelzer, & Turner, 2015; Cuddy et al., 2012; Vanstone & Cuddy, 2010; Baird & Samson, 2009; Cuddy & Duffin, 2005). For example, patients with the expected gray matter atrophy profile associated with Alzheimer's disease had impairments in semantic memory but intact musical memory, which was supported by a network that included bilateral supplementary motor cortex and left anterior superior temporal cortex (Slattery et al., 2019). The patterns of atrophy (whether from acute damage or a degenerative disorder) that selectively affect some memory systems more than others lend support to the idea that there are separate networks for musical and verbal memory.

Although there is general agreement that musical memories are spared in neurodegenerative disorders, a recent meta-analysis found little consistency in the brain areas involved in memory for music (Freitas et al., 2018). Generally, the recognition of familiar music appears to rely on a fronto-temporal network (Slattery et al., 2019; Agustus et al., 2018; Jacobsen et al., 2015; Sikka, Cuddy, Johnsrude, & Vanstone, 2015; Herholz, Halpern, & Zatorre, 2012; Groussard et al., 2009; Plailly, Tillmann, & Royet, 2007; Halpern & Zatorre, 1999) along with SMAs (Slattery et al., 2019; Agustus et al., 2018; Herholz et al., 2012; Pereira et al., 2011; Peretz et al., 2009) and basal ganglia structures (Agustus et al., 2018; Sikka et al., 2015; Pereira et al., 2011). However, no two studies are in agreement about the brain areas necessary for musical memory.

In this study, we investigated which neural networks are responsible for the processing of music and language and how they are affected by memory for the stimuli. To our knowledge, only one other study has investigated how memory for music and memory for language interact (Saito et al., 2012). In that study, participants listened to familiar and unfamiliar children's songs (recreated on voice-synthesizing software) while undergoing an O^{15} PET activation scan. The analysis uncovered separate networks for the retrieval of familiar music and language. Familiar music stimuli recruited the right middle temporal sulcus and bilateral temporo-occipital cortices, and familiar language stimuli recruited the left fusiform gyrus and the left inferior occipital gyrus, adding to the disagreement in the literature regarding which areas are involved in musical memory.

To expand on previous work, the stimuli in the current study were designed to be as similar to what is heard "in the real world" as possible and to manipulate the presence or absence of music and language. Previous experiments have differed in the type of stimuli used to probe musical memory with some stimuli containing music with language (e.g., songs with lyrics) and others containing music without language (e.g., classical musical excerpts), but it is unknown how the presence or absence of language influences memory for music. To understand how music and language interact during learning and memory formation, four stimulus conditions were created: (1) whole music (music and words together), (2) instrumental music without words, (3) a capella (sung words, no instruments), and (4) spoken words. This allowed for an assessment of whether the effect of memory differed based on music or language content and whether stimuli with more information (i.e., music AND language, rather than each independently) would be remembered differently.

Participants completed a strict training paradigm to control for their knowledge of the stimuli. This training process is in contrast to studies that compare novel music to music made familiar over a lifetime. Relying on lifetime exposure to a piece of music makes it impossible to control familiarity across participants and, therefore, to untangle the differences between memory for the music and the autobiographical memories linked to the music. In addition, learning music shortly before a retrieval phase (e.g., Esfahani-Bayerl et al., 2019; Alonso et al., 2016) may not accurately produce the level of familiarity that occurs "naturally" with repeated exposure over longer time periods. Therefore, the training paradigm in this study was designed to mimic exposure to music

over time while carefully monitoring the amount of exposure. Participants listened to the music through a specialized music player that tracked the number of times a participant heard each stimulus. The training process created an objective measure of familiarity with the stimuli and allowed for comparisons between stimuli that differed only in the participant's degree of familiarity. By carefully manipulating the presence of language and the degree of familiarity, this study's aim was to bring some clarity to the disagreement in the field and to better understand the relationship between the neural networks responsible for music and language and how they interact with memory abilities.

METHODS

Ethics

Ethics approval for this project was granted by the Health Sciences Research Ethics Board at The University of Western Ontario (#100606, #114263).

Participants

Twenty-six neurologically healthy, English-speaking participants (14 female) aged 18–39 (mean = 24 years) were recruited at The University of Western Ontario. All participants had completed at least some postsecondary education, and nine participants had completed some postgraduate education. Using the Goldsmith's Musical Sophistication Index (Müllensiefen, Gingras, Musil, & Stewart, 2014), 17 participants reported having formal musical training (1–10 years, mean = 4.5 years), but, at the time of testing, only nine of them played instruments regularly. Seven participants were fluent in a second language. All participants reported listening to music regularly (average 1.5 hr per day) via a phone, computer, or car radio. No further data on the diversity of the participants' backgrounds were collected, and therefore, no comment can be made on whether the results from the current sample of participants are generalizable to a more diverse sample.

Two individuals withdrew from the study following the first scan session, and data from four individuals were not included in the analysis because the average scores on the behavioral memory tests (lyric modification and melody memory test—details below) were lower than 70% correct. fMRI data from 20 individuals were included in the analysis.

Stimuli

Stimuli were similar to those regularly encountered in the real world, and the presence of language and music was manipulated. Stimuli were created from the lyrics and music of eight different songs written and recorded by one of the authors (A. M. O.) between 1997 and 2006 for an amateur rock band based in Cambridge, United Kingdom. Thus, all stimuli were completely novel to the Canadian participants. The original songs were all written in a similar style, and instrumentation included a lead singer, bass, drums, guitar, string instruments, and backing vocals, each recorded on separate tracks. Stimuli from the band's original repertoire were selected based on having male vocals only (over some that included female vocals). All stimuli were recorded using the same equipment directly to digital hard drive using the Sonar software (by Cakewalk) and a ShureSM58 microphone. Where the same instruments appear across stimuli (violin, cello, drums, guitar, etc.), the same physical instruments were used.

Four conditions were created by modifying the original eight songs to include only certain tracks: (1) whole music (music and words together in a fully intact version of each song), (2) instrumental music without words (all vocal parts were removed, leaving just the nonvocal instrumentation), (3) a capella (all nonvocal instrumentation was removed leaving just the lead and backing vocals), and (4) spoken words (the lyrics of each song were rerecorded in spoken form by the original lead singer to have a similar length, tempo, and emotional intonation as their original song counterparts). There were two different stimuli for each condition, and none of the original songs were used for more than one condition.

The stimuli varied in length from 3:00 to 4:03 min. However, during the fMRI scan sessions, participants heard only 10-sec clips taken from the stimuli. Equal numbers of clips were taken from the beginning, middle, and end of the stimuli. Clips were taken from both verses and chorus and were chosen such that musical phrases were not interrupted within the clip. Each of the 10-sec clips was normalized to equate their perceived loudness using the *Audacity* software (Audacity Team, 2020). Further details regarding the acoustic characteristics were determined using the Praat software (Boersma & Weenink, 2018) and can be found in Table 1. During the training period, participants listened to half of the stimuli (four stimuli, one per condition) via an on-line audio player. The full stimuli, as well as the 10-sec clips, can be found in the Supplementary Information.

There were a total of four learning categories of stimuli: "To be learned" refers to the novel stimuli heard in the first scanning session that the participant subsequently listened to over the training period; "not to be learned" refers to the novel stimuli heard in the first scanning session that the participant *did not* listen to over the training period; "learned" refers to the stimuli heard in the second scanning session that the participant listened to over the training period; and "not learned" refers to the stimuli in the second scanning session that the participant *did not* listen to over the training period. The "to be learned" and "learned" stimuli were identical for each participant, as were the "not to be learned" and "not

Table 1. A Summary of the Average Acoustic Characteristics (as Determined by Praat Software) for the 10-Sec Stimulus Clips Used in This Experiment

	Average Pitch (Hz)	Average Pitch Range (Hz)	Average Harmonicity (dB)	Average Tempo (bpm)
a capella music	229.79	398.14	11.82	135
instrumental music	146.71	524.86	4.36	162
spoken word	105.97	528.55	8.62	–
whole music	138.27	622.37	3.92	156

learned" stimuli. The sets of stimuli that were learned were counterbalanced across participants: Half the participants familiarized with one half of the stimuli; the other half of the participants familiarized with the other half of the stimuli (Groups A and B; see Table 2).

Procedure

Participants completed two fMRI scans that were separated by a stimulus training period (14–29 days; mean = 20 days). During both scans, participants passively listened to the stimuli. During the training period, participants listened to the stimuli via an on-line audio player (designed in-lab) that tracked the number of times each stimulus was played. Participants were asked to listen to the stimuli at least 5 times per week. To ensure participants were engaged while listening, the player presented a simple question about the stimulus (e.g., "*Were there lyrics present in the previous song?*") at random between stimuli. A response was required to move to the next stimulus. Participants

were encouraged to incorporate the music into their everyday lives (i.e., to listen while cooking or driving).

Behavioral Familiarity Tasks

Participants came to the laboratory every few days to complete a total of four behavioral testing sessions between their two scans. In each session, participants listened to the stimuli and completed a series of behavioral tasks. Each session lasted less than 1 hr, and the behavioral tasks described below were distributed across sessions.

We created two tests to track participants' familiarity with the stimuli. The first was a lyric modification task that visually presented participants with two sentences. One sentence was a lyric taken directly from the participant's training stimuli group, and the other sentence was a modified version of the same lyric. Participants indicated which of the sentences was correct. The correct and incorrect lyric pairs were tested for validity before the study to ensure that modified lyrics were chosen at least

Table 2. Notes to Description of How the Stimuli Were Counterbalanced across Participants and How the Behavioral Tasks Were Designed to Probe Learning of Both Music and Language

Eight different stimuli	1	2	3	4	5	6	7	8
Conditions	A Capella	Instrumental Music	Spoken Word	Whole Music	A Capella	Instrumental Music	Spoken Word	Whole Music
Group A participants (n = 11)	Learned				Not learned			
Group B participants (n = 9)	Not learned				Learned			
Lyric modification task	x		x	x	x		x	x
	Group A task 21 lyric pairs (probing learning in stimuli with language)				Group B task 29 lyric pairs (probing learning in stimuli with language)			
Melody recognition task	x	x		x	x	x		x
	Group A task (probing learning in stimuli with music)				Group B task (probing learning in stimuli with music)			

equally as often as original lyrics in naive listeners. Two versions of the task were created to probe learning of the lyrics in the stimuli learned by Groups A and B (see Table 2). Because of there being a larger number of word repetitions in the Group A stimuli, more lyric pairs were included in the lyric modification task for Group B to account for the larger number of unique words in the Group B stimuli.

Before the first scan session, participants were tested on the full set of lyric pairs, but as they were not yet familiar with any of the stimuli, they were asked to indicate which lyric they believed was *most likely* to come from a real song. During the behavioral sessions, participants were presented with a randomly generated subset of 10 lyric pairs to track learning progress. Participants were tested on the full set of lyric pairs again after the second scan session. Only conditions that contained words (whole music, a capella, and spoken) were tested (see Table 2).

The second test of familiarity was a melody recognition task. After the second scan *only*, participants listened to 23 pairs of 2-sec clips taken from the stimuli. Three to four clips were taken from each stimulus, and none of the clips contained any lyrics. Melodic information was extracted from the a capella stimuli using the Praat program (Boersma & Weenink, 2018). During the task, participants were presented with one clip taken from a stimulus the participant trained on and a second clip from a stimulus the participant did not train on (in a randomized order). Participants indicated which of the two clips was most familiar to them. Only conditions that contained melodies (whole music, a capella, and instrumental) were tested (see Table 2).

To ensure the familiar stimuli were truly familiar, any participant who scored an average of 70% correct or less across the two tasks was excluded from further analyses.

Preference Ratings

In each laboratory session and after the second scan, participants rated from 1 to 5 how much they liked the stimuli, allowing us to track changes in preference with increased familiarity.

On-Line Task Verification

A separate cohort of 32 participants completed an on-line version of the same lyric modification task described previously *without* training on the stimuli. The on-line study was used to determine whether an increase in scores from the first to last sessions could be attributed to training with the stimuli or simply because of exposure to the lyric modification task. The task was completed via on-line surveys that were e-mailed to participants at time intervals that mimicked the original study. Participants were asked to complete the surveys within 24 hr of receiving the e-mail.

During the first session, participants completed the full set of lyric pairs and then listened to all eight stimuli

once. Participants were randomly assigned to one of two groups to match the counterbalanced training groups from the original study. In each of the subsequent four sessions, participants completed a short survey of 10 lyric pairs from the stimuli in their "learning" group. In the final session, participants listened to all eight stimuli for a second time and completed the full set of lyric pairs mimicking the order of events from the original study. Participants only listened to the stimuli in the first and last on-line sessions and did not have access to the stimuli in the interim period. No fMRI data were collected from participants completing the on-line study.

fMRI Acquisition and Analyses

Imaging was conducted at the Robarts Research Institute on a Siemens Magnetom 7 Tesla scanner with a 32-channel head coil. Functional scans were acquired with 54 slices per volume (repetition time = 1.25 sec; echo time = 20 msec; flip angle = 35°; field of view = 220 × 220 mm; voxel size = 2.5 mm^3). The two scan sessions (before and after the training period) were identical and included two 12-min functional runs. Participants heard ten 10-sec clips from each of the eight stimuli (80 clips total) that were randomized across the two runs (40 clips in each run). Half of these clips were "to be learned" in the first session and "learned" in the second session, whereas the other half were "not to be learned" in the first session and "not learned" in the second session. Between functional runs in the first session only, a whole-head anatomical scan was acquired (repetition time = 6 sec; echo time = 2.69 msec; field of view = 240 × 240 mm; voxel size = 0.75 mm^3; 208 slices).

Data were processed using SPM12. Data were corrected for motion and coregistered to the participant's structural image. Images were normalized to Montreal Neurological Institute space, and smoothing was done with a Gaussian kernel of 8-mm FWHM. Subject-specific first-level models combined data from all four runs (two from the first session and two from the second session) and included epochs representing each of the 10-sec stimulus clips convolved by the canonical hemodynamic response function. Covariates of no interest, representing six motion parameters (x, y, z, translation and rotation) were also included. Serial correlations were accounted for using an autoregressive model, and low-frequency noise was removed with a high-pass filter of 128 sec. Contrast images from single-participant models were created for each of the eight stimuli versus rest in each session for a total of 16 contrast images per subject. The 16 contrasts were then entered into a second-level full-factorial model for a group-level analysis. The second-level model factors were session (eight stimuli in the first session, eight stimuli in the second), stimulus type (two stimuli for each of the four types in each session), and learning condition (four sessions, one stimuli = to be learned; four sessions, one stimuli = not to be learned;

Table 3. Areas Identified by More than One Study in the Meta-Analysis by Freitas et al. (2018) as Being Involved in Memory for Music

Area Name	Harvard–Oxford Division
Insular cortex	2
Superior frontal gyrus	3
Middle frontal gyrus	4
IFG (triangularis)	5
IFG (opercularis)	6
Precentral gyrus	7
Superior temporal lobe (anterior)	9
Superior temporal lobe (posterior)	10
Middle temporal gyrus (anterior)	11
Middle temporal gyrus (posterior)	12
Cingulate gyrus (anterior)	29
Cingulate gyrus (posterior)	30

ROI templates from the Harvard–Oxford atlas were used.

four sessions, two stimuli = learned; four sessions two stimuli = not learned).

Contrasts probing differences in stimulus type, learning condition, and session were generated at the group level. To probe differences in memory, a 2 (session) × 2 (learning) ANOVA was conducted within each of the four stimulus types. For each stimulus type, the main effect of session was calculated by comparing Session 1 and Session 2 (e.g., [a capella not to be learned + a capella to be learned] vs. [a capella not learned + a capella learned]). These t contrasts were calculated in both directions (Session 1 > Session 2; Session 2 > Session 1). The main effect of learning in each stimulus type was calculated both across sessions and within Session 2. Learning across session compared Session 1 to be learned with Session 2 learned (e.g., a capella to be learned vs. a capella learned) and learning within session compared Session 2 learned with Session 2 not learned (e.g., a capella learned vs. a capella not learned). These t contrasts were generated in both directions. Finally, interaction contrasts were generated within each stimulus type: ([to be learned – not to be learned] – [learned – not learned]). To probe differences in language, pairwise contrasts were created between each of the four stimulus categories (a capella vs. instrumental, a capella vs. spoken, a capella vs. whole, instrumental vs. spoken, instrumental vs. whole, and spoken vs. whole). These t contrasts were generated separately in each session. Bayesian statistics were implemented using built-in SPM12 functions. To test for more subtle changes in the patterns of brain activation between conditions, we used a multivariate representational similarity analysis (RSA). The RSA was implemented using The RSAToolbox (Diedrichsen, Provost, & Zareamoghaddam, 2016) in 12 bilateral Harvard–Oxford

defined ROIs (atlas distributed with the FMRIB Software Library software package fsl.fmrib.ox.ac.uk/fsl/). The ROIs were selected on the basis of a meta-analysis that identified these areas as being involved in memory for music (Freitas et al., 2018; see Table 3).

RESULTS

Participant Training

Participants listened to the stimuli an average of 13 times (from 6 to 20 listens) over an average of 20 days (from 14 to 29 days).

Behavioral Familiarity Tasks

Participants significantly improved on the lyric modification task over the training period (see Figure 1). During the first session, participants scored an average of 36% correct, which was significantly lower than the average 82% correct score during the final session, $t(34) = -12.3, p < .011, d = 2.62$ (with three participants scoring over 90%). The below-chance performance of individuals during the first session is because of how the lyric modification pairs were created. The lyric pairs were designed such that the modified lyrics were chosen at least as often as the original lyrics in naive listeners. Often, to make the meaning of the new lyric seem *plausible*, we made choices that made the lyric seem more *likely* than
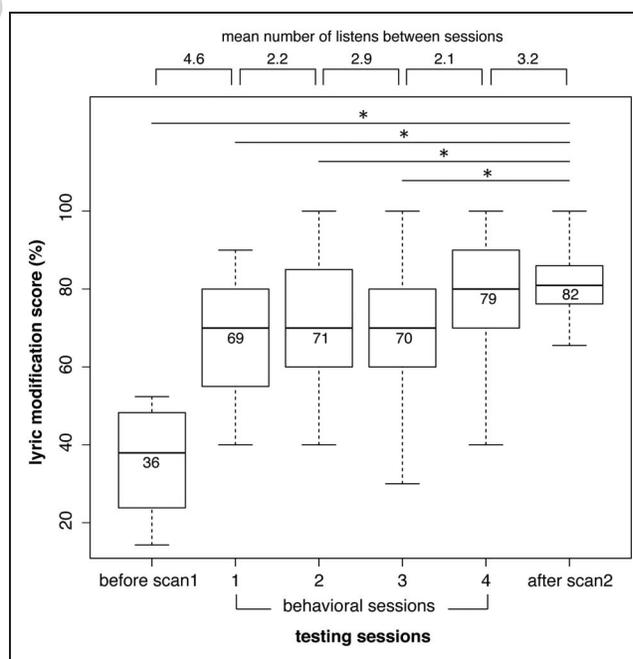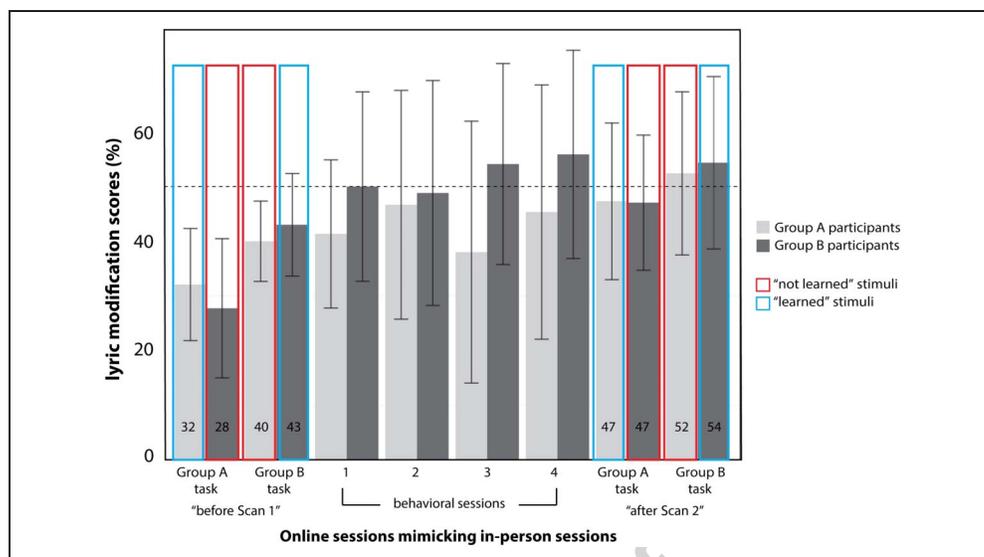

Figure 1. Scores on the lyric modification task across all sessions averaged across the two learning groups. Boxplots show the average, minimum, and maximum correct scores within each session. The average score in each session is listed within each box. The average number of times participants listened to the stimuli between each session is listed above the figure. Significant differences in scores between sessions are shown (*$p < .05$).

Figure 2. Thirty-two participants completed an on-line version of the lyric modification task *without training* on the stimuli to determine whether the increase in scores seen in the original experiment was because of exposure to the task itself rather than learning of the stimuli. Error bars represent plus/minus 1 *SD*. Red and blue boxes highlight scores representing the "not learned" or "learned" stimuli, respectively. Average scores are listed on the "before scan 1" and "after scan 2" bars directly.

the original (when songwriters write lyrics, word choices are often not made based on plausibility, but on intended meaning whether plausible or not). Thus, as a result of fine-tuning our lyric pairs, there were a number of modified lyrics that were initially chosen more often than the original.

There was no difference in average scores between the two learning groups in the final session (A: 80% vs. B: 85%; $t(15) = -0.66, p = .52, d = 0.3$). Scores from the first behavioral testing session 69% correct, $t(32) = -3.2, p = .003, d = 0.86$), the second behavioral testing session (71% correct, $t(30) = -2.5, p = .02, d = 0.83$), and the third behavioral testing session (70% correct, $t(30) = -2.6, p = .02, d = 0.71$) were also significantly lower than the scores recorded during the final testing session. The scores recorded in the fourth behavioral testing session did not significantly differ from the scores recorded during the final session (79% correct, $t(30) = -0.7, p = .49, d = 0.18$). Scores on the lyric modification task did not differ between the three conditions tested (spoken, whole, and a capella).

In the on-line task verification study, there was a significant increase from the first session (average 35%) to the last session (average 50%; $t(121) = -6.30, p < .001, d = 0.84$; see Figure 2). However, when comparing the "not learned" stimuli (i.e., Group A doing the Group B task and Group B doing the Group A task—outlined in red) and the "learned" stimuli (i.e., Group A doing the Group A task and Group B doing the Group B task—outlined in blue), there was no interaction between learning and session.

A 2 × 2 mixed measures ANOVA with session (first, last) and experiment (fMRI, on-line) was performed to compare learning in the fMRI and on-line experiments. There was a main effect of experiment, with participants performing better in the fMRI experiment, $F(1, 164) = 45.5, p < .001, \eta^2 = .12$, and a main effect of session, with participants performing better in the final session, $F(1, 164) = 124.7, p < .001, \eta^2 = .33$. However, these must

be interpreted in light of a statistically significant interaction between session and experiment, $F(1, 164) = 46.2, p < .001, \eta^2 = .12$ (Figure 3).

Examination of mean scores in Figure 3 clearly shows that, although both groups of participants began at similar levels, the group who actively trained on the stimuli improved significantly more over time than the group who were merely tested repeatedly over the same timeframe. Thus, for the first session, scores from the on-line study did not differ from the fMRI study (36% vs. 36%, $t(28) = 0.04, p = .97, d = 0.01$), whereas, for the final session, scores from the on-line study were significantly lower than the fMRI study (50% vs. 82%, $t(47) = -11.2, p < .001, d = 2.34$; Figure 2). These results indicate that, although exposure to the task alone did increase scores, it was not enough to explain the substantial improvement in the experimental group who trained on the stimuli over the same time period.
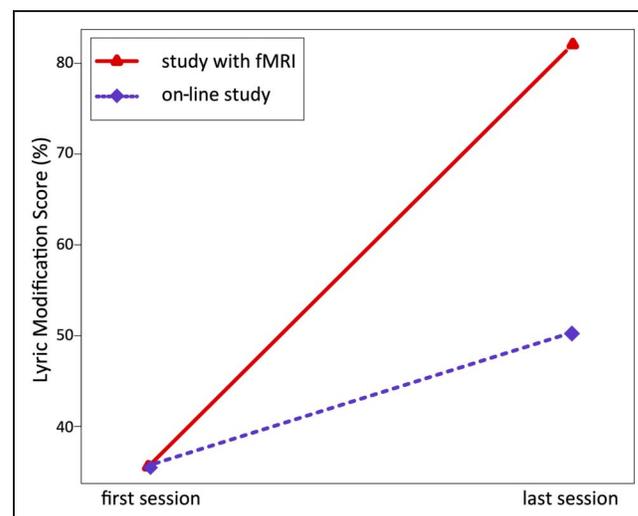

Figure 3. Lyric modification scores from the first and last sessions of the fMRI and on-line experiments.

Table 4. Preference Ratings for the Stimuli Averaged Over All Testing Sessions

Stimulus Type	Average Preference Rating
Whole	3.55 ± 1.07
Instrumental	3.64 ± 1.14
A capella	2.57 ± 1.06
Spoken	2.10 ± 1.11

Participants scored an average of 92% ($SD = 6.4$) on the melody memory task completed during the second session, indicating that they were near ceiling at recognizing the melodies of the stimuli they heard during the training period.

Before the training period, participants' preference ratings across all stimuli were an average of 2.9/5. After the training period, there was no change in participants' preference ratings of the stimuli, $t(38) = -0.17, p = .87, d = 0.07$. This was true for all types of stimuli. Average preference ratings for each stimulus type over all testing sessions can be found in Table 4. There was a significant difference in preference based on the type of stimulus, $F(3, 796) = 93.82, p < .001, \eta^2 = .26$. On average, the participants liked the whole stimuli significantly more than the a capella, $t(398) = 9.19, p < .001, d = 1.02$, or the spoken stimuli, $t(398) = 13.29, p < .001, d = 1.04$. Preferences for the instrumental stimuli did not differ from the whole stimuli, $t(396) = 0.77, p = 0.44, d = 0.08$. Participants also preferred the instrumental stimuli over the a capella, $t(396) = 9.64, p < .001, d = 0.97$, and spoken stimuli, $t(398) = 13.61, p < .001, d = 0.95$. Participants preferred the a capella stimuli over the spoken stimuli, $t(398) = 4.32, p < .001, d = 0.33$.
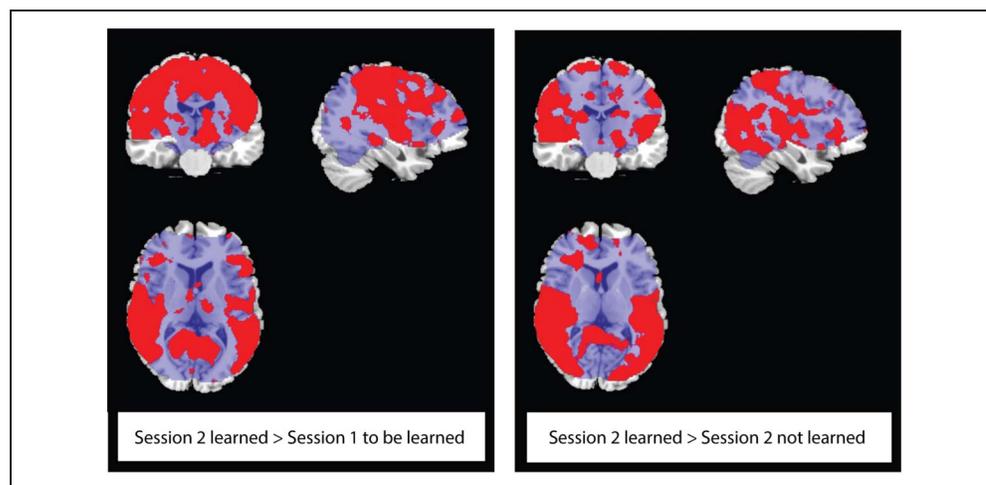
fMRI Results

Memory

For each subject, all first-level contrasts for the 16 stimuli (8 different stimuli across 2 sessions) were entered into a second-level full-factorial model using SPM12. For each of the four stimulus types (whole, instrumental, a capella, and spoken), a 2 × 2 ANOVA was performed to test for significant effects of learning and session. Eight contrasts (two for each stimulus type) were created probing average stimulus differences in Session 1 versus Session 2 (Session 1 > Session 2; Session 2 > Session 1). Eight contrasts (two for each stimulus type) were created probing learning across session (Session 1 to be learned > Session 2 learned; Session 2 learned > Session 1 to be learned). Eight contrasts (two for each stimulus type) were created probing learning within Session 2 (Session 2 learned > Session 2 not learned; Session 2 not learned > Session 2 learned). Four interaction contrasts (one for each stimulus type) were created probing the interaction between session and learning ([to be learned – not to be learned] – [learned – not learned]). There were no significant main effects or interactions between learning and session for any of the four stimulus types in any of the contrasts listed above (all: $t < 3.7, p > .6$).

We further investigated the pairwise comparisons' null results using the built-in Bayesian statistics toolbox in SPM12 (default Cohen's $d = 1.0$). The toolbox reports evidence in support of the alternative hypothesis (i.e., evidence in support of a significant difference between conditions). However, we were interested in the opposite evidence in support of the null hypothesis (i.e., evidence in support of *no* difference between conditions). Therefore, a custom script was created to extract areas that supported the null hypothesis at a Bayes Factor level of 1/50 (very strong evidence for the null; Stefan, Gronau, Schönbrodt, & Wagenmakers, 2019). Applying Bayesian statistics to the two pairwise contrasts described above showed that activity levels in several brain areas were very likely not to significantly differ between conditions. Thus, the areas highlighted in Figure 4 are statistically 50× more likely to not differ (i.e., in support of the null) than they are to differ.

Finally, an RSA probed for differences in voxel activity patterns generated by the learned and not learned

Figure 4. Bayesian statistical results for two contrasts at BF = 1/50. Left: Session 2 learned stimuli > Session 1 to be learned stimuli (identical stimuli differing in familiarity only). Right: Session 2 learned stimuli > Session 2 not learned stimuli (different stimuli). Bayesian statistics were only applied to areas of the brain in which data were acquired in all 20 participants (shown in blue). Thus, for example, the lack of statistics in the area of the basal ganglia reflects the fact that not all 20 participants contributed data in that area. Crosshairs are at $x = 35, y = -18, z = 9$.

Session 2 learned > Session 1 to be learned

Session 2 learned > Session 2 not learned

stimulus conditions across the two 12-min functional runs in the second session. Within the 12 ROIs, ß-weights for each individual and each stimulus type were extracted and spatially prewhitened using an estimate of the overall noise-covariance matrix (Walther et al., 2016) resulting in the remaining noise in each voxel being approximately uncorrelated with the noise in other voxels (Diedrichsen & Kriegeskorte, 2017; Diedrichsen et al., 2016). We then quantified the difference between the prewhitened patterns of activity using a "crossnobis estimator" (an unbiased method of determining the distance between patterns, as the estimator's average will be zero if two
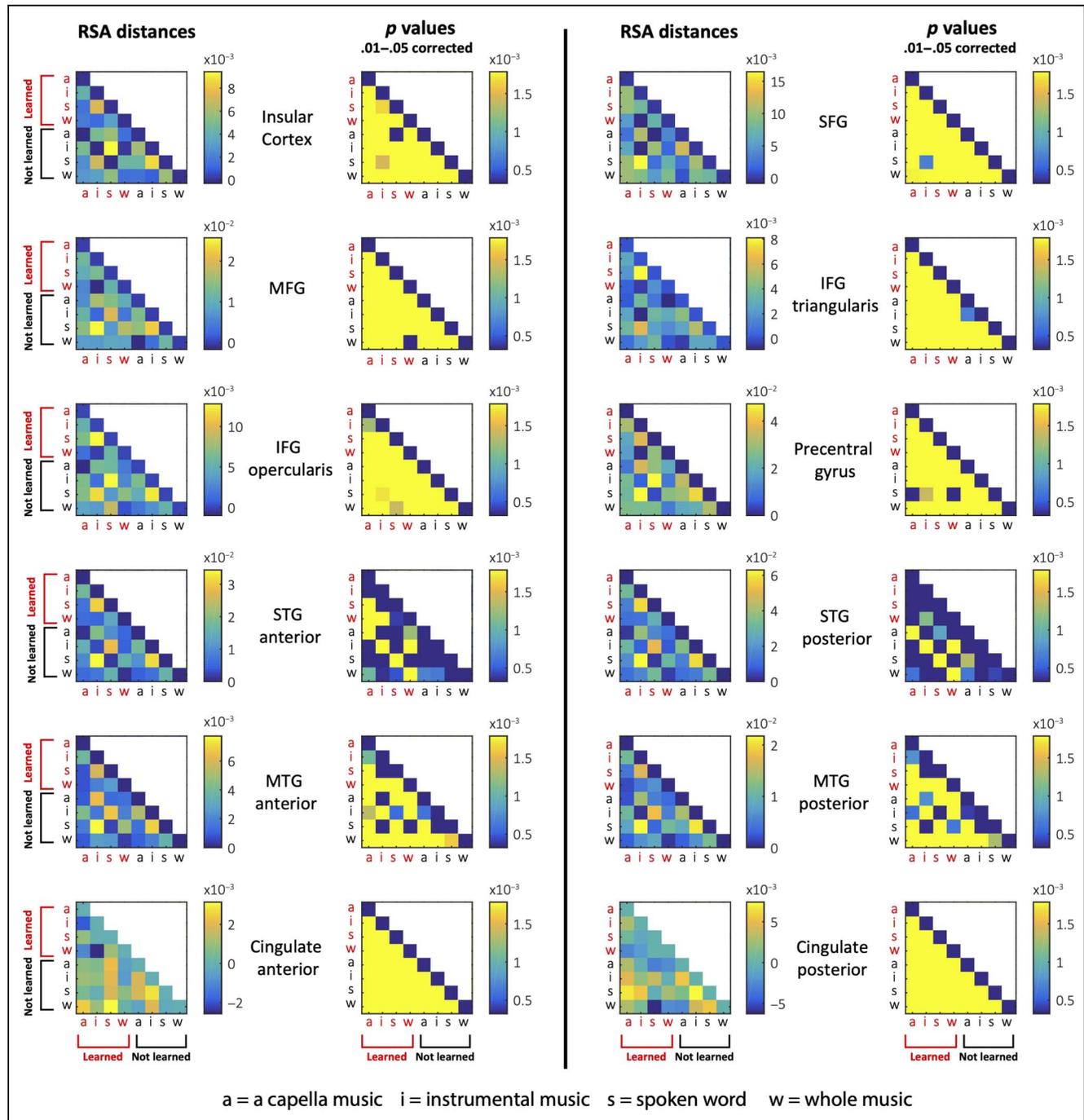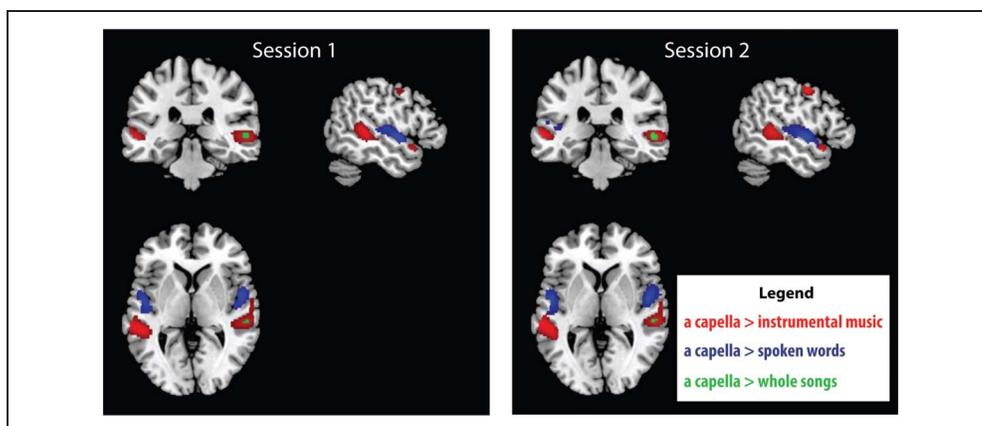
Figure 5. Distances between the patterns for the "learned" (in red letters) and "not learned" (in black letters) stimuli (from the second session), in 12 ROIs. The matrices in the left columns indicate the distance between patterns generated by each pair of stimuli. The matrices in the right columns are the corresponding *p* values (Bonferroni corrected) from a one-tailed *t* test of those distances across all 20 participants. A yellow color square in the right matrix indicates that the *t* test on the corresponding square in the left matrix is not significant at a corrected $p \geq .05$ level. Any dark blue squares in the right matrices indicate significance at a level of $p \leq .01$. The diagonal of the lower left 4×4 corner of each matrix contains the direct comparisons between the learned and not learned conditions (i.e., a cappella learned vs. a cappella not learned, instrumental learned vs. instrumental not learned, spoken learned vs. spoken not learned, and whole learned vs. whole not learned).

Table 5. Results from Binary Contrasts between All Stimulus Categories in Both Sessions

Contrast	Region	Session 1		Session 2	
		p FDR < .05	*Coordinates (x, y, z)*	*p FDR < .05*	*Coordinates (x, y, z)*
A capella > instrumental	L posterior temporal (superior gyrus)	< .001	−52, −38, 2	< .001	−54, −36, 2
	R posterior temporal (superior and middle gyri)	< .001	58, −16, −6	< .001	60, −26, −2
	L precentral gyrus	.068	−52, −2, 46	.006	−50, −2, 46
A capella > spoken	L planum polare	< .001	−48, −2, −4	< .001	−48, −4 ,−4
	R planum polare	< .001	52, −2, 0	< .001	50, −4, 0
A capella > whole	R posterior temporal (superior and middle gyri)	.003	60, −18, −4	.005	56, −28, −2
Instrumental > a capella	–	–	–	–	–
Instrumental > spoken	L planum polare	.001	−46, −6, −4	< .001	−46, −4, −6
	R planum polare	< .001	48, 2, −6	< .001	48, 2, −8
	R planum temporale	.003	62, −26, 16	–	–
	L angular gyrus	–	–	.048	−58, −56, 38
Instrumental > whole	–	–	–	–	–
Spoken > a capella	–	–	–	–	–
Spoken > instrumental	L posterior temporal (superior gyrus)	< .001	−52, −38, 2	< .001	−54, −36, 2
	R posterior temporal (middle gyrus)	< .001	48, −26, −4	< .001	54, −30, −2
	L inferior frontal (pars triangularis)	.014	−52, 22, 16	–	–
	Temporal pole	–	–	.018	−50, 14, −16
Spoken > whole	L posterior temporal (middle gyrus)	< .001	−54, −34, −2	–	–
	R posterior temporal (middle gyrus)	< .001	52, −22, −6	< .001	52, −30, −2
Whole > a capella	–	–	–	–	–
Whole > instrumental	L posterior temporal (superior and middle gyri)	< .001	−52, −38, 2	< .001	−54, −36, 2
	R anterior temporal (superior gyrus)	.006	60, −4, −8	–	–
Whole > spoken	L planum polare	< .001	−46, −6, −4	< .001	−46, −2, −4
	R planum polare	< .001	50, −2, 0	< .001	48, 2, −8

Significant clusters at FDR correction < 0.05 are listed. Missing results indicate no significant clusters were found.

Figure 6. The left shows results from the first scan, and the right shows results from the second scan. Three contrasts are shown: a capella > instrumental music (red), a capella > spoken words (blue), a capella > whole songs (green). Crosshairs are placed at $x = -51$, $y = -30$, $z = 3$.

patterns differ only by noise). The resulting distances between the two patterns were plotted as a representational dissimilarity matrix (see Figure 5). We compared the four "learned" and the four "not learned" stimuli. Negative distances can be interpreted as zero distance, or no evidence of dissimilarity between stimulus categories. Mathematically, the negative distance is derived from whether or not the distance between the two conditions was consistent across the runs. In this case, there were two runs included in the analysis and the negative distance is likely a result of the distance vectors between the conditions of interest ("learned" and "not learned") being in slightly different space. When the inner product of the vectors from the two runs is taken, the resultant vector length is less than zero. There were no systematic RSA differences between the "learned" and the "not learned" stimuli across the 12 ROIs. There were RSA differences between the different stimulus categories in temporal auditory areas: bilateral anterior and posterior middle temporal gyrus and superior temporal gyrus.

Language

Given the lack of main effects or interactions resulting from familiarity with the stimuli and the fact that the stimuli presented in the two scanning sessions were identical, the second session was treated as a replication of the first to investigate the reliability of the stimulus type differences between sessions. Using the same full-factorial model as described previously that contained all 16 first-level contrasts from each participant (8 different songs across 2 sessions), pairwise contrasts between each of the four stimulus categories (a capella, instrumental, spoken, and whole) were calculated within each session. Brain areas with significant activity differences between stimulus types are listed in Table 5. Results from the two sessions are shown side by side to show the consistency between the two independently collected sessions. The statistical contrasts were only calculated in brain areas that contained data from all 20 participants (shown in blue in Figure 4).

A capella stimuli generated significantly more activity than other stimulus categories in auditory areas (posterior superior and middle temporal gyri, planum polare) and left motor areas (precentral gyrus) in both sessions (see Figure 6).

Instrumental stimuli generated significantly more activity than spoken stimuli in bilateral auditory cortices (planum polare) in both sessions and in the left angular gyrus in the second session only (see Figure 7). Instrumental stimuli did not result in more activity than the a capella or whole stimuli in either session.

Spoken stimuli generated significantly more activity than instrumental and whole stimuli in auditory cortices

Figure 7. The left shows results from the first scan, and the right shows results from the second scan. One contrast is shown: instrumental music > spoken words. Crosshairs are placed at $x = -53$, $y = -8$, $z = 2$.
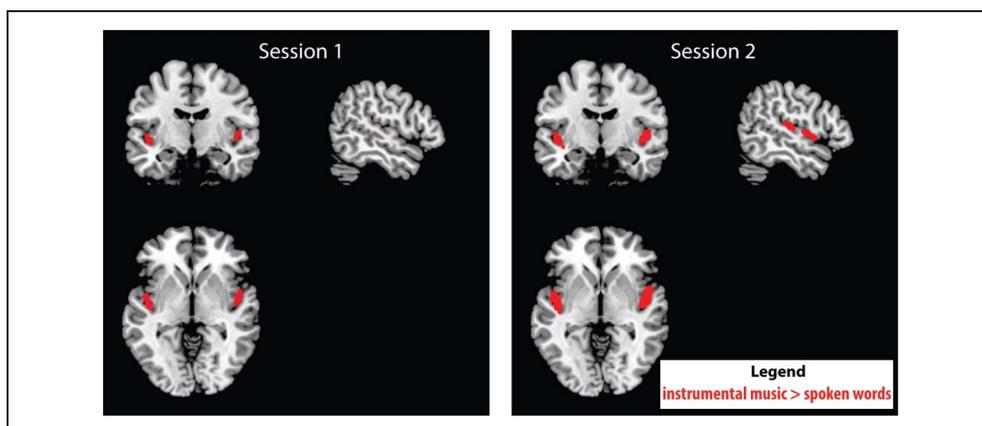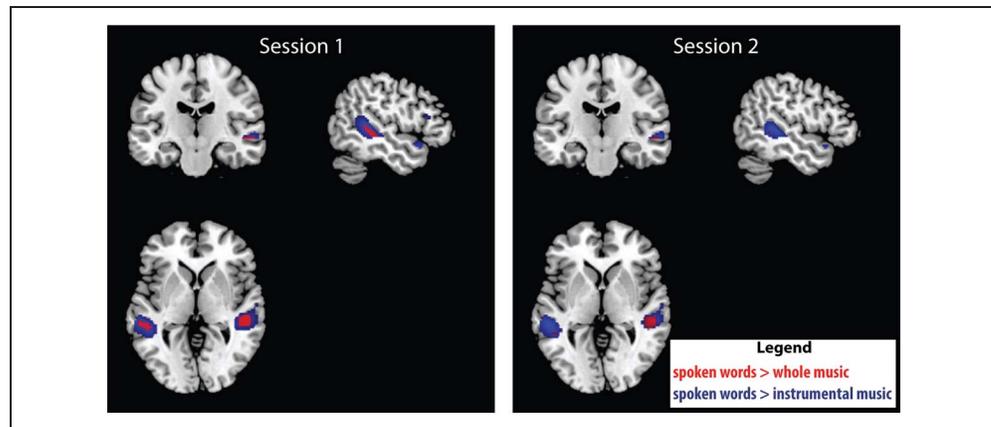
Figure 8. The left shows results from the first scan, and the right shows results from the second scan. Two contrasts are shown: spoken words > whole music (red), spoken words > instrumental music (blue). Crosshairs are placed at $x = -50$, $y = -18$, $z = 1$.

(posterior superior and middle gyri) in both sessions. Spoken stimuli also resulted in significantly more activity than instrumental in the left inferior frontal area (pars triangularis) in the first session and in the temporal pole in the second session (see Figure 8).

Whole stimuli resulted in significantly more activity than instrumental and spoken stimuli in bilateral auditory cortices areas (posterior superior and middle temporal gyri, planum polare) in both sessions. Whole stimuli did not produce more activity than a capella stimuli in either session.

DISCUSSION

In the current study, we set out to understand the relationship between the neural networks responsible for music and language abilities and how they are influenced by memory. The four stimulus conditions allowed us to compare four combinations of music and language, namely, whole music, instrumental music, a capella, and spoken word. Using a novel methodological approach consisting of a strictly controlled training paradigm, we isolated natural exposure to music while controlling for autobiographical memory confounds. We monitored the number of times each stimulus was heard by participants, creating an objective measure of familiarity.

To track familiarity with the language component of the stimuli, participants identified correct lyrics in a forced choice paradigm. Scores on the lyric modification task improved during the training period (i.e., between the two fMRI scans), providing objective verification that the stimuli had, in fact, become more familiar over time. Moreover, the on-line follow-up study confirmed that performance improvement was the result of training exposure, rather than simply repeated exposure to the task itself. To track learning of the melodic component of the stimuli, participants identified familiar melodies in a forced choice task. This task was completed after the second fMRI scan session. Performance was near ceiling. The results from these two tests indicate that the participants became familiar with both the language and the musical components of the stimuli over the training period.

fMRI responses to the four stimulus types differed. This result can be seen in the RSA in the anterior and posterior middle and superior temporal gyri as well as in the general linear model contrast analyses. Because familiarity did not significantly affect activation to the stimuli, we treated the two scanning sessions separately in the general linear model contrast analysis, replicating the comparisons between different stimulus types in each session. The three conditions that contained music (a capella, instrumental, and whole music) activated the bilateral planum polare more than the spoken condition did. The planum polare activity did not differ between the three musical conditions. The planum polare, along with the inferior frontal gyrus (IFG), has been shown to play a role in processing language and musical syntax, with increasing stimulus complexity resulting in more activation (Merrill et al., 2012; Brown, Martinez, & Parsons, 2006; Constable et al., 2004; Bookheimer, 2002; Griffiths, Büchel, Frackowiak, & Patterson, 1998). It may be that the lack of planum polare activation differences between the three music conditions is because of the highly controlled nature of the stimuli. The stimuli that contained music were written by the same individual and likely did not differ in musical complexity; however, as these stimuli contained more musical information than the spoken condition, planum polare activation may have been greater as a result.

The left IFG was significantly more active in the spoken than in the instrumental condition. Despite the IFG's known involvement in musical and language syntax processing (Kunert, Willems, Casasanto, Patel, & Hagoort, 2015; Merrill et al., 2012; Brown et al., 2006), the significant difference in activation was only seen in the one comparison (i.e., not seen in the a capella-instrumental or whole-instrumental comparisons). The activation difference between the spoken and instrumental conditions was in the opposite direction to that seen in the planum polare (where more activity to instrumental music than to spoken stimuli was found). A similar activation pattern was reported in another passive listening task that also directly compared music and language stimuli: greater activation in planum polare to stimuli with melodic pitch

information than stimuli without pitch information, and greater activation in IFG to stimuli with language information over stimuli without language information (Merrill et al., 2012). In the current study, the need to process language syntax information was higher for the spoken stimuli than for the instrumental stimuli and is likely the reason for the difference in IFG activation. However, this reasoning does not explain why no differences were seen between either the a capella or the whole stimuli, which also contained language information, over the instrumental stimuli. It is possible that this pattern of results can be explained by an interaction between language and musical syntax processing. For example, Kunert et al. (2015) found that IFG activation only occurs when both the music and language components of the stimuli are syntactically challenging. In the current study, the two stimuli that were most different from each other in terms of both musical and language syntax were those in the spoken and instrumental conditions leading to the statistically significant difference in IFG activation.

All three conditions that contained language (a capella, spoken, and whole) significantly activated bilateral posterior superior and middle temporal gyri more than the instrumental condition. These areas have been previously identified in passive listening tasks involving speech and nonspeech stimuli (Tie et al., 2014; Tremblay, Baroni, & Hasson, 2013) and are "voice-selective" (Fecteau, Armony, Joanette, & Belin, 2004; Belin, Zatorre, Lafaille, Ahad, & Pike, 2000). During passive listening to vocal (words, phrases, sentences, etc.) and nonvocal sounds (machine noises, nature sounds etc.), areas along the superior and middle temporal gyri are more active for stimuli with vocalizations than without vocalizations (Belin et al., 2000; Fecteau et al., 2004). In a direct comparison between speech and musical instruments, the posterior portions of the superior and middle temporal gyri were more active to human voice than to nonvocal sounds (Bethmann & Brechmann, 2014). In the current study, activity did not differ across the three different conditions involving the voice, confirming that these areas are generally active in response to the human voice.

The current experiment did not include a vocal no-language condition (e.g., humming), nor a language non-vocal condition (e.g., computerized language) to dissociate the presence of language from the presences of vocal sounds. Although the differences in the bilateral posterior superior and middle temporal gyri may be attributable to the presence of vocalizations, rather than language, the IFG and the planum polare are not known to be "voice-selective" (Belin et al., 2000). Therefore, the activation differences in the IFG and planum polare are likely because of differences in language and syntax processing between stimulus categories (Merrill et al., 2012).

It is possible that the differences in acoustic characteristics across stimuli may have influenced the results. To mitigate any stimulus-specific effects, two stimuli were included in each of the four stimulus conditions, and the stimuli were counterbalanced across learning groups such that all stimuli were included in each of the four learning groups (to be learned, not to be learned, learned, not learned). Therefore, it is unlikely that a single stimulus could drive the differences between conditions, but we cannot definitively state that condition differences were completely uninfluenced by acoustic differences between conditions.

Although the behavioral task results confirmed that participants were more familiar with the stimuli during the second fMRI scan than the first, there were no corresponding neural changes associated with this behavioral improvement. The Bayesian statistics support this result at a Bayes Factor level of 1/50 (i.e., very strong evidence to support no difference between the "learned" and "not learned" conditions). An RSA (a multivariate approach that takes into account the subtle pattern variations in brain activation to different stimuli) also showed no difference between the "learned" and "not learned" stimuli. One issue to consider is that the imaging data were collected at a 7-Tesla magnetic field strength. In high-strength magnets, there can be signal loss in some brain areas. Here, signal dropout occurred in anterior temporal areas and basal ganglia across the majority of participants. However, primary auditory area activity was preserved. Effects of familiarity may be observed at lower field strengths in the areas where signal was lost in the current study. It is also possible that the null results found in the brain areas with consistent data across all participants were because of undetected true effects and an increase in stimulus trials or sample size could increase power.

The lack of difference between novel and familiar music is in contrast with the results of other studies (e.g., Freitas et al., 2018; Herholz et al., 2012; Halpern & Zatorre, 1999) and is likely related to procedural differences. In most previous studies, participants listened to stimuli that they already knew, such as children's songs or folksongs (e.g., Alonso et al., 2016; Schaal, Javadi, Halpern, Pollok, & Banissy, 2015; Herholz et al., 2012; Saito et al., 2012), popular music from the radio charts (e.g., Jacobsen et al., 2015; Pereira et al., 2011), or music supplied by the participants (e.g., El Haj, Fasotti, & Allain, 2012), and rated their familiarity with the stimuli. Using well-known music does not control for the amount of exposure to the stimuli, but it does reflect the way individuals generally learn and become familiar with music "in the real world." In contrast, in the current study, participants learned novel stimuli by listening both in a "sterile" laboratory environment, as well as out of the laboratory via an on-line music player and the number of exposures to the stimuli gave an objective measure of familiarity. Although participants were encouraged to incorporate the music into their everyday lives (i.e., to listen while cooking or driving), few participants reported having done so. Participants listened to the stimuli an average of 13 times over the course of the study, which is likely fewer times than a well-known song is heard over a lifetime. For example,

the number one pop song in Canada is played over 5,600 times per week across all Canadian radio stations (World Airplay Radio Monitor: Real-Time Radio Tracking, 2012). A person may encounter that song hundreds of times over the course of their life. Therefore, although the participants learned the current stimuli, they did not learn them to the same level as songs "in the real world."

In addition, the behavioral tests to probe music familiarity may have inflated measures of how well participants learned the stimuli because they relied on recognition rather than recollection memory. Recognition memory requires a more "shallow" encoding of the stimuli being remembered than does recollection (Mandler, 2008). Although participants performed well on the recognition tests, we expect that if participants had been asked to sing the stimuli, their recollection would be worse than if they sang a well-known song like, for example, a Christmas carol. However, assessing recall of musical information in participants, especially those who are not musicians, is difficult to do accurately as it requires the separation of deficits in recall from deficits in musical ability. For example, poor singing of a song could be caused by poor song recollection or poor singing ability. The reliance on recognition memory tests in the current study may have led us to believe that the differences in memory between the novel and familiar stimuli were more profound than in actuality.

When people listen to music in the real world, that listening is often connected with many other aspects of life (i.e., people, places, or experiences), making it difficult to separate the specific musical memory response in the brain from the neural response to the autobiographical memories evoked by that music. One study examined the overlap between musical and autobiographical memories evoked by music (Janata, 2009) and identified areas in which activation correlated with the degree of autobiographical salience above and beyond the degree of familiarity. These areas were located in pFC in bilateral superior frontal gyrus (Brodmann's areas 8 and 9) and in left IFG (BA 45). These same regions have been identified as involved in musical memory in studies that presented individuals with well-known music (Groussard et al., 2009; Klostermann, Loui, & Shimamura, 2009; Plailly et al., 2007). It is possible that the areas previously attributed to musical memory are in fact activated because of the autobiographical memories triggered by the music, rather than by memory for the music itself. In this study, the carefully controlled way in which participants learned the stimuli did not allow participants to create the autobiographical memories they would have if the music was learned "naturally." Without such memories, the differences between the brain activity patterns for the "learned" and "not learned" stimuli are presumably reduced, limiting the ability to detect differences based on familiarity with the music alone.

Participant's lack of preference for the stimuli, exacerbated by the way the stimuli were created, likely also contributed to the lack of familiarity results in the fMRI data. To create the different stimulus categories while controlling for as many features as possible, we deconstructed whole songs into their component parts. Although spoken word, instrumental music, and a capella music are all genres of their own, the way they were created in this study was not representative of these genres. For example, to create the instrumental and a capella music stimuli, we extracted only the specific lines of interest (instrumental or sung voice) from original whole songs. This process resulted in music that was not representative of instrumental and a capella music because each line was musically simpler and less interesting to listen to on its own, as, originally, they were intended to be listened to as part of a larger whole. Similarly, the spoken word stimuli were created by recording song lyrics as spoken words. Although similar to poetry, the lyrics were not written to be experienced without music and participants informally reported that the repetitive nature of the lyrics was not pleasant to listen to as prose. In comparison, the whole stimuli were not modified. Participants' preferences mirrored the amount each stimulus was modified: Participants preferred the stimuli that were modified the least over those modified the most. This pattern was not related to participants' memory for the stimuli as measured by the behavioral tasks. Interestingly, participants' enjoyment of the stimuli did not increase with exposure, as would be expected by the mere exposure effect (an increase in preference as a result of repeated exposure; Zajonc, 2001). Although there were no directly negative events associated with listening to the stimuli, the requirement to listen to the stimuli daily and the association with the laboratory environment may have been enough to override any mild positive reactions a participant may have had from the repeated stimulus exposure, resulting in no change in their preference ratings. Using stimuli intended to be experienced as poetry, a capella, or instrumental music (rather than deconstructing whole stimuli) may have increased participants' preference and memory for the stimuli, as preferred music is better remembered than nonpreferred music (Stalinski & Schellenberg, 2013; Samson, Dellacherie, & Platel, 2009; Eschrich, Münte, & Altenmüller, 2008). However, using such existing stimuli would not have allowed control of similarity across the stimulus types (all stimuli were written by the same individual, with similar instrumentation, the same voice across all stimuli, and from a similar rock genre).

This study isolated memory for music from the confounding factor of autobiographical memory by asking participants to train on highly controlled novel stimuli. As a result, we have come to understand a number of key components that are necessary for musical memory. The way individuals engage with music is important for creating a memory for that music. The degree of engagement during the learning of music, driven by preference or autobiographical memories associated with the music,

may speak to why there is such a disagreement in the literature about the areas involved in musical memory as it is difficult to control for participants' engagement while maintaining a natural learning process. Further investigations into how musical memory, emotional engagement, and language processing are related may be key to understanding what makes memory for music so unique and robust in the presence of neurodegenerative disorders such as Alzheimer's disease.

Reprint requests should be sent to Avital Sternin, Brain and Mind Institute, University of Western Ontario, N6A5B7 London, Ontario, Canada, or via e-mail: asternin@uwo.ca.

Author Contributions

Funding Information

Supporting Information

The full stimuli and the 10-sec clips used during this experiment can be found at the following url: https://owenlab.uwo.ca/research/research_tools.html.

Diversity in Citation Practices

A retrospective analysis of the citations in every article published in this journal from 2010 to 2020 has revealed a persistent pattern of gender imbalance: Although the proportions of authorship teams (categorized by estimated gender identification of first author/last author) publishing in the *Journal of Cognitive Neuroscience (JoCN)* during this period were M(an)/M = .408, W(oman)/M = .335, M/W = .108, and W/W = .149, the comparable proportions for the articles that these authorship teams cited were M/M = .579, W/M = .243, M/W = .102, and W/W = .076 (Fulvio et al., *JoCN*, 33:1, pp. 3–7). Consequently, *JoCN* encourages all authors to consider gender balance explicitly when selecting which articles to cite and gives them the opportunity to report their article's gender citation balance. The authors of this article report its proportions of citations by gender category to be as follows: M/M = .372, W/M = .267, M/W = .085, and W/W = .277.

REFERENCES

Agustus, J. L., Golden, H. L., Callaghan, M. F., Bond, R. L., Benhamou, E., Hailstone, J. C., et al. (2018). Melody processing characterizes functional neuroanatomy in the aging brain. *Frontiers in Neuroscience*, *12*, 815. https://doi.org/10.3389/fnins.2018.00815, PubMed: 30524219

Alonso, I., Davachi, L., Valabrègue, R., Lambrecq, V., Dupont, S., & Samson, S. (2016). Neural correlates of binding lyrics and melodies for the encoding of new songs. *Neuroimage*, *127*, 333–345. https://doi.org/10.1016/j.neuroimage.2015.12.018, PubMed: 26706449

Audacity Team. (2020). *Audacity(R): Free audio editor and recorder*. https://audacityteam.org/.

Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusia. *Brain*, *125*, 238–251. https://doi.org/10.1093/brain/awf028, PubMed: 11844725

Baird, A., & Samson, S. (2009). Memory for music in Alzheimer's disease: Unforgettable? *Neuropsychology Review*, *19*, 85–101. https://doi.org/10.1007/s11065-009-9085-2, PubMed: 19214750

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, *403*, 309–312. https://doi.org/10.1038/35002078, PubMed: 10659849

Bethmann, A., & Brechmann, A. (2014). On the definition and interpretation of voice selective activation in the temporal cortex. *Frontiers in Human Neuroscience*, *8*, 499. https://doi.org/10.3389/fnhum.2014.00499, PubMed: 25071527

Boersma, P., & Weenink, D. (2018). *Praat: Doing phonetics by computer* (6.0.37) [Computer software]. Celemony. www.praat.org.

Bookheimer, S. (2002). Functional MRI of language: New approaches to understanding the cortical organization of semantic processing. *Annual Review of Neuroscience*, *25*, 151–188. https://doi.org/10.1146/annurev.neuro.25.112701.142946, PubMed: 12052907

Brown, S., Martinez, M. J., & Parsons, L. M. (2006). Music and language side by side in the brain: A PET study of the generation of melodies and sentences. *European Journal of Neuroscience*, *23*, 2791–2803. https://doi.org/10.1111/j.1460-9568.2006.04785.x, PubMed: 16817882

Constable, R. T., Pugh, K. R., Berroya, E., Mencl, W. E., Westerveld, M., Ni, W., & Shankweiler, D. (2004). Sentence complexity and input modality effects in sentence comprehension: An fMRI study. *Neuroimage*, *22*, 11–21. https://doi.org/10.1016/j.neuroimage.2004.01.001, PubMed: 15109993

Cuddy, L. L., & Duffin, J. (2005). Music, memory, and Alzheimer's disease: Is music recognition spared in dementia, and how can it be assessed? *Medical Hypotheses*, *64*, 229–235. https://doi.org/10.1016/j.mehy.2004.09.005, PubMed: 15607545

Cuddy, L. L., Duffin, J. M., Gill, S. S., Brown, C. L., Sikka, R., & Vanstone, A. D. (2012). Memory for melodies and lyrics in

Alzheimer's disease. *Music Perception: An Interdisciplinary Journal*, *29*, 479–491. https://doi.org/10.1525/mp.2012.29.5 .479, PubMed: 15607545

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review*, *3*, 422–433. https://doi.org/10.3758 /BF03214546, PubMed: 24213976

Deutsch, D., Henthorn, T., & Lapidis, R. (2011). Illusory transformation from speech to song. *Journal of the Acoustical Society of America*, *129*, 2245–2252. https://doi .org/10.1121/1.3562174, PubMed: 21476679

Diedrichsen, J., & Kriegeskorte, N. (2017). Representational models: A common framework for understanding encoding, pattern-component, and representational-similarity analysis. *PLoS Computational Biology*, *13*, e1005508. https://doi.org /10.1371/journal.pcbi.1005508, PubMed: 28437426

Diedrichsen, J., Provost, S., & Zareamoghaddam, H. (2016). On the distribution of cross-validated Mahalanobis distances. *ArXiv:1607.01371 [Stat]*. https://arxiv.org/abs/1607.01371.

El Haj, M., Fasotti, L., & Allain, P. (2012). The involuntary nature of music-evoked autobiographical memories in Alzheimer's disease. *Consciousness and Cognition*, *21*, 238–246. https://doi.org/10.1016/j.concog.2011.12.005, PubMed: 22265372

Eschrich, S., Münte, T. F., & Altenmüller, E. O. (2008). Unforgettable film music: The role of emotion in episodic long-term memory for music. *BMC Neuroscience*, *9*, 48. https://doi.org/10.1186/1471-2202-9-48, PubMed: 18505596

Esfahani-Bayerl, N., Finke, C., Kopp, U., Moon, D.-U., & Ploner, C. J. (2019). Musical memory and hippocampus revisited: Evidence from a musical layperson with highly selective hippocampal damage. *Cortex*, *119*, 519–527. https://doi.org /10.1016/j.cortex.2018.12.023, PubMed: 30795831

Fadiga, L., Craighero, L., & D'Ausilio, A. (2009). Broca's area in language, action, and music. *Annals of the New York Academy of Sciences*, *1169*, 448–458. https://doi.org/10.1111 /j.1749-6632.2009.04582.x, PubMed: 19673823

Fecteau, S., Armony, J. L., Joanette, Y., & Belin, P. (2004). Is voice processing species-specific in human auditory cortex? *An fMRI study. Neuroimage*, *23*, 840–848. https://doi.org/10 .1016/j.neuroimage.2004.09.019, PubMed: 15528084

Finke, C., Esfahani, N. E., & Ploner, C. J. (2012). Preservation of musical memory in an amnesic professional cellist. *Current Biology*, *22*, R591–R592. https://doi.org/10.1016/j.cub.2012.05 .041, PubMed: 22877775

Freitas, C., Manzato, E., Burini, A., Taylor, M. J., Lerch, J. P., & Anagnostou, E. (2018). Neural correlates of familiarity in music listening: A systematic review and a neuroimaging meta-analysis. *Frontiers in Neuroscience*, *12*, 686. https://doi .org/10.3389/fnins.2018.00686, PubMed: 30344470

Gordon, R. L., Schön, D., Magne, C., Astésano, C., & Besson, M. (2010). Words and melody are intertwined in perception of sung words: EEG and behavioral evidence. *PLoS One*, *5*, e9889. https://doi.org/10.1371/journal.pone.0009889, PubMed: 20360991

Griffiths, T. D. (1997). Spatial and temporal auditory processing deficits following right hemisphere infarction. A psychophysical study. *Brain*, *120*, 785–794. https://doi.org/10.1093/brain/120.5 .785, PubMed: 9183249

Griffiths, T. D., Büchel, C., Frackowiak, R. S. J., & Patterson, R. D. (1998). Analysis of temporal structure in sound by the human brain. *Nature Neuroscience*, *1*, 422–427. https://doi.org/10 .1038/1637, PubMed: 10196534

Groussard, M., Viader, F., Landeau, B., Desgranges, B., Eustache, F., & Platel, H. (2009). Neural correlates underlying musical semantic memory. *Annals of the New York Academy of Sciences*, *1169*, 278–281. https://doi.org/10.1111/j.1749-6632.2009 .2009.04784.x, PubMed: 19673793

Halpern, A. R., & Zatorre, R. J. (1999). When that tune runs through your head: A PET investigation of auditory imagery for familiar melodies. *Cerebral Cortex*, *9*, 697–704. https://doi .org/10.1093/cercor/9.7.697, PubMed: 10554992

Herholz, S., Halpern, A., & Zatorre, R. (2012). Neuronal correlates of perception, imagery, and memory for familiar tunes. *Journal of Cognitive Neuroscience*, *24*, 1382–1397. https://doi.org/10.1162/jocn_a_00216, PubMed: 22360595

Hymers, M., Prendergast, G., Liu, C., Schulze, A., Young, M. L., Wastling, S. J., et al. (2015). Neural mechanisms underlying song and speech perception can be differentiated using an illusory percept. *Neuroimage*, *108*, 225–233. https://doi.org /10.1016/j.neuroimage.2014.12.010, PubMed: 25512041

Jackendoff, R. (2009). Parallels and nonparallels between language and music. *Music Perception: An Interdisciplinary Journal*, *26*, 195–204. https://doi.org/10.1525/mp.2009.26.3.195

Jackendoff, R., & Lerdahl, F. (2006). The capacity for music: What is it, and what's special about it? *Cognition*, *100*, 33–72. https://doi.org/10.1016/j.cognition.2005.11.005, PubMed: 16384553

Jacobsen, J., Fritz, T., Stelzer, J., & Turner, R. (2015). Why musical memory can be preserved in advanced Alzheimer's disease. *Brain*, 1–13. https://doi.org/10.1093/awv148

Janata, P. (2009). The neural architecture of music-evoked autobiographical memories. *Cerebral Cortex*, *19*, 2579–2594. https://doi.org/10.1093/cercor/bhp008, PubMed: 19240137

Klostermann, E. C., Loui, P., & Shimamura, A. P. (2009). Activation of right parietal cortex during memory retrieval of nonlinguistic auditory stimuli. *Cognitive, Affective, & Behavioral Neuroscience*, *9*, 242–248. https://doi.org/10 .3758/CABN.9.3.242, PubMed: 19679760

Koelsch, S., Gunter, T., Cramon, D., Zysset, S., Lohmann, G., & Friederici, A. (2002). Bach speaks: A cortical "language-network" serves the processing of music. *Neuroimage*, *17*, 956–966. https://doi.org/10.1006/nimg.2002.1154, PubMed: 12377169

Kunert, R., Willems, R. M., Casasanto, D., Patel, A. D., & Hagoort, P. (2015). Music and language syntax interact in Broca's area: An fMRI study. *PLoS One*, *10*, e0141069. https:// doi.org/10.1371/journal.pone.0141069, PubMed: 26536026

Mandler, G. (2008). Familiarity breeds attempts: A critical review of dual-process theories of recognition. *Perspectives on Psychological Science*, *3*, 390–399. https://doi.org/10.1111 /j.1745-6924.2008.00087.x, PubMed: 26158957

Merrill, J., Sammler, D., Bangert, M., Goldhahn, D., Lohmann, G., Turner, R., et al. (2012). Perception of words and pitch patterns in song and speech. *Frontiers in Psychology*, *3*, 76. https://doi.org/10.3389/fpsyg.2012.00076, PubMed: 22457659

Metz-Lutz, M.-N., & Dahl, E. (1984). Analysis of word comprehension in a case of pure word deafness. *Brain and Language*, *23*, 13–25. https://doi.org/10.1016/0093-934X(84) 90002-6

Mohanan, M. K. P. (1982). *Lexical phonology* (PhD Thesis). Massachusetts Institute of Technology.

Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLoS One*, *9*, e89642. https://doi.org/10.1371/journal.pone.0089642, PubMed: 24586929

Norman-Haignere, S. V., Feather, J., Brunner, P., Ritaccio, A., McDermott, J. H., Schalk, G., et al. (2019). Intracranial recordings from human auditory cortex reveal a neural population selective for musical song [Preprint]. *Neuroscience*. https://doi.org/10.1101/696161

Patel, A. D. (2008). *Music, language, and the brain*. Oxford University Press.

Pereira, C. S., Teixeira, J., Figueiredo, P., Xavier, J., Castro, S. L., & Brattico, E. (2011). Music and emotions in the brain:

Familiarity matters. *PLoS One*, *6*, e27241. https://doi.org/10
.1371/journal.pone.0027241, PubMed: 22110619

Peretz, I. (1996). Can we lose memory for music? A case of
music agnosia in a nonmusician. *Journal of Cognitive
Neuroscience*, *8*, 481–496. https://doi.org/10.1162/jocn.1996
.8.6.481, PubMed: 23961980

Peretz, I., & Coltheart, M. (2003). Modularity of music processing.
Nature Neuroscience, *6*, 688–691. https://doi.org/10.1038
/nn1083, PubMed: 12830160

Peretz, I., Gosselin, N., Belin, P., Zatorre, R. J., Plailly, J., &
Tillmann, B. (2009). Music lexical networks: The cortical
organization of music recognition. *Annals of the New York
Academy of Sciences*, *1169*, 256–265. https://doi.org/10.1111
/j.1749-6632.2009.04557.x, PubMed: 19673789

Peretz, I., Kolinsky, R., Tramo, M., Labrecque, R., Hublet, C.,
Demeurisse, G., et al. (1994). Functional dissociations
following bilateral lesions of auditory cortex. *Brain*, *117*,
1283–1301. https://doi.org/10.1093/brain/117.6.1283,
PubMed: 7820566

Peretz, I., & Zatorre, R. J. (2005). Brain organization for music
processing. *Annual Review of Psychology*, *56*, 89–114.
https://doi.org/10.1146/annurev.psych.56.091103.070225,
PubMed: 15709930

Piccirilli, M., Sciarma, T., & Luzzi, S. (2000). Modularity of music:
Evidence from a case of pure amusia. *Journal of Neurology,
Neurosurgery & Psychiatry*, *69*, 541–545. https://doi.org/10
.1136/jnnp.69.4.541, PubMed: 10990522

Plailly, J., Tillmann, B., & Royet, J.-P. (2007). The feeling of
familiarity of music and odors: The same neural signature?
Cerebral Cortex, *17*, 2650–2658. https://doi.org/10.1093
/cercor/bhl173, PubMed: 17289777

Saito, Y., Ishii, K., Sakuma, N., Kawasaki, K., Oda, K., &
Mizusawa, H. (2012). Neural substrates for semantic memory
of familiar songs: Is there an interface between lyrics and
melodies? *PLoS One*, *7*, e46354. https://doi.org/10.1371
/journal.pone.0046354, PubMed: 23029492

Samson, S., Dellacherie, D., & Platel, H. (2009). Emotional
power of music in patients with memory disorders: Clinical
implications of cognitive neuroscience. *Annals of the New
York Academy of Sciences*, *1169*, 245–255. https://doi.org/10
.1111/j.1749-6632.2009.04555.x, PubMed: 19673788

Schaal, N. K., Javadi, A.-H., Halpern, A. R., Pollok, B., & Banissy,
M. J. (2015). Right parietal cortex mediates recognition
memory for melodies. *European Journal of Neuroscience*,
42, 1660–1666. https://doi.org/10.1111/ejn.12943, PubMed:
25959620

Schön, D., Gordon, R., Campagne, A., Magne, C., Astésano, C.,
Anton, J.-L., et al. (2010). Similar cerebral networks in
language. *music and song perception. Neuroimage*, *51*,
450–461. https://doi.org/10.1016/j.neuroimage.2010.02.023,
PubMed: 20156575

Sikka, R., Cuddy, L. L., Johnsrude, I. S., & Vanstone, A. D.
(2015). An fMRI comparison of neural activity associated
with recognition of familiar melodies in younger and older
adults. *Frontiers in Neuroscience*, *9*, 1–10. https://doi.org/10
.3389/fnins.2015.00356, PubMed: 26500480

Slattery, C. F., Agustus, J. L., Paterson, R. W., McCallion, O.,
Foulkes, A. J. M., Macpherson, K., et al. (2019). The functional
neuroanatomy of musical memory in Alzheimer's disease.
Cortex, *115*, 357–370. https://doi.org/10.1016/j.cortex.2019.02
.003, PubMed: 30846199

Stalinski, S. M., & Schellenberg, E. G. (2013). Listeners
remember music they like. *Journal of Experimental
Psychology: Learning, Memory, and Cognition*, *39*, 700–716.
https://doi.org/10.1037/a0029671, PubMed: 22905936

Stefan, A. M., Gronau, Q. F., Schönbrodt, F. D., &
Wagenmakers, E.-J. (2019). A tutorial on Bayes factor design
analysis using an informed prior. *Behavior Research
Methods*, *51*, 1042–1058. https://doi.org/10.3758/s13428-018
-01189-8, PubMed: 30719688

Takahashi, N., Kawamura, M., Shinotou, H., Hirayama, K., Kaga,
K., & Shindo, M. (1992). Pure word deafness due to left
hemisphere damage. *Cortex*, *28*, 295–303. https://doi.org/10
.1016/S0010-9452(13)80056-1, PubMed: 1499314

Tie, Y., Rigolo, L., Norton, I. H., Huang, R. Y., Wu, W., Orringer,
D., et al. (2014). Defining language networks from resting-
state fMRI for surgical planning—A feasibility study. *Human
Brain Mapping*, *35*, 1018–1030. https://doi.org/10.1002/hbm
.22231, PubMed: 23288627

Tierney, A., Dick, F., Deutsch, D., & Sereno, M. (2013). Speech
versus song: Multiple pitch-sensitive areas revealed by a naturally
occurring musical illusion. *Cerebral Cortex*, *23*, 249–254. https://
doi.org/10.1093/cercor/bhs003, PubMed: 22314043

Tremblay, P., Baroni, M., & Hasson, U. (2013). Processing of
speech and non-speech sounds in the supratemporal plane:
Auditory input preference does not predict sensitivity to
statistical structure. *Neuroimage*, *66*, 318–332. https://doi.org
/10.1016/j.neuroimage.2012.10.055, PubMed: 23116815

Vanstone, A. D., & Cuddy, L. L. (2010). Musical memory in
Alzheimer disease. *Neuropsychology, Development, and
Cognition, Section B: Aging, Neuropsychology and
Cognition*, *17*, 108–128. https://doi.org/10.1080
/13825580903042676, PubMed: 19657762

Walther, A., Nili, H., Ejaz, N., Alink, A., Kriegeskorte, N., &
Diedrichsen, J. (2016). Reliability of dissimilarity measures for
multi-voxel pattern analysis. *Neuroimage*, *137*, 188–200.
https://doi.org/10.1016/j.neuroimage.2015.12.012, PubMed:
26707889

World Airplay Radio Monitor: Real-Time Radio Tracking. (2012).
https://warmmusic.net.

Yaqub, B. A., Gascon, G. G., Nosha, M. A., & Whitaker, H.
(1988). Pure word deafness (acquired verbal auditory
agnosia) in an Arabic speaking patient. *Brain*, *111*, 457–466.
https://doi.org/10.1093/brain/111.2.457, PubMed: 3378145

Zajonc, R. B. (2001). Mere exposure: A gateway to the
subliminal. *Current Directions in Psychological Science*,
10, 224–228. https://doi.org/10.1111/1467-8721.00154

Zatorre, R. J., & Gandour, J. T. (2008). Neural specializations for
speech and pitch: Moving beyond the dichotomies.
*Philosophical Transactions of the Royal Society of London,
Series B: Biological Sciences*, *363*, 1087–1104. https://doi.org
/10.1098/rstb.2007.2161, PubMed: 17890188