



The effects of short-term musical training on the neural processing of speech-in-noise in older adults

David Fleming^{f,a,*}, Sylvie Belleville^{c,d}, Isabelle Peretz^{b,d}, Greg West^d, Benjamin Rich Zendel^{a,b,c,e,*}

^a Faculty of Medicine, Memorial University of Newfoundland, St. John's, Newfoundland and Labrador A1B 3V6, Canada

^b International Laboratory for Brain, Music, and Sound Research, Montréal, Québec H3C 3J7, Canada

^c Centre de Recherche, Institut Universitaire de Gériatrie de Montréal (CRIUGM), Montréal, Québec H3W1W4, Canada

^d Département de Psychologie, Université de Montréal, Québec H3C 3J7, Canada

^e Aging Research Centre – Newfoundland and Labrador, Memorial University of Newfoundland, Corner Brook, Newfoundland and Labrador A2H5G4, Canada

^f PPLE College, University of Amsterdam, Amsterdam 1018WB, the Netherlands

ARTICLE INFO

Keywords:

Aging
Music-training
fMRI
Speech perception

ABSTRACT

Experienced musicians outperform non-musicians in understanding speech-in-noise (SPIN). The benefits of lifelong musicianship endure into older age, where musicians experience smaller declines in their ability to understand speech in noisy environments. However, it is presently unknown whether commencing musical training in old age can also counteract age-related decline in speech perception, and whether such training induces changes in neural processing of speech. Here, we recruited older adult non-musicians and assigned them to receive a short course of piano or videogame training, or no training. Participants completed two sessions of functional Magnetic Resonance Imaging where they performed a SPIN task prior to and following training. While we found no direct benefit of musical training upon SPIN perception, an exploratory Region of Interest analysis revealed increased cortical responses to speech in left Middle Frontal and Supramarginal Gyri which correlated with changes in SPIN task performance in the group which received music training. These results suggest that short-term musical training in older adults may enhance neural encoding of speech, with the potential to reduce age-related decline in speech perception.

1. Introduction

Musicianship is associated with benefits in understanding speech-in-noise (SPIN) (for a review, see Coffey, Mogilever, & Zatorre, 2017). The neural encoding of speech is enhanced in musicians (Kraus & Chandrasekaran, 2010; Parbery-Clark, Strait, & Kraus, 2011; Zendel, Tremblay, Belleville, & Peretz, 2015), who show increased activity in, and interactions between, auditory and motor cortical regions during SPIN perception (Du & Zatorre, 2017), consistent with the idea that speech processing is fine-tuned by learning a musical instrument because of shared neural architecture which supports both abilities (i.e., recruitment of similar auditory sub-cortical and cortical pathways, and sensori-motor connections by both processes; Patel, 2011, 2012). The ability to understand SPIN declines with age, but this decline is offset by musicianship: older musicians are better able to understand SPIN relative to age-matched non-musicians (Parbery-Clark, Anderson, Hittner, & Kraus, 2012; Parbery-Clark, Strait, Anderson, Hittner, & Kraus, 2011; Zendel & Alain, 2012).

Studies comparing musician and non-musicians typically focus on groups of highly-trained musicians with many years' experience. These cross-sectional designs have two shortcomings. Firstly, musicians and non-musicians are self-selected and any differences in speech processing could be pre-existing and not due to training. Secondly, these studies select musicians with many years of training (often, in the case of older musicians, most of a lifetime). This leaves open the critical question if short-term musical training in older adults could lead to improvements in speech perception. If so, music training may be useful as a form of auditory rehabilitation for older adults.

Training studies in younger participants are promising. In children, SPIN perception improvements have been reported after 2 years of musical training, but not after 1 year (Slater et al., 2015). Furthermore, short-term music training in children and adolescents leads to improvements in behavioural performance on linguistic tasks and enhancements in neural responses to speech (e.g., modulated amplitude of auditory cortical event-related potentials, and improved consistency and earlier timing of auditory brainstem responses reflecting speech

* Corresponding authors at: PPLE College, University of Amsterdam, Amsterdam 1018WB, the Netherlands (D. Fleming). Faculty of Medicine, Memorial University of Newfoundland, 230 Elizabeth Ave, St. John's, NL A1B 3V6, Canada (B.R. Zendel).

E-mail addresses: dfleming@mun.ca (D. Fleming), bzendel@mun.ca (B.R. Zendel).

<https://doi.org/10.1016/j.bandc.2019.103592>

Received 11 July 2018; Received in revised form 29 July 2019; Accepted 1 August 2019

0278-2626/ © 2019 Elsevier Inc. All rights reserved.

encoding) (Moreno & Besson, 2006; Moreno, Lee, Janus, & Bialystok, 2015; Moreno, Marques, & Santos, 2009; Tierney, Krizman, & Kraus, 2015; Tierney, Krizman, Skoe, Johnston, & Kraus, 2013) with training duration ranging from 4 to 8 weeks to 1 year in children (Fujioka, Ross, Kakigi, Pantev, & Trainor, 2006; Moreno & Besson, 2006; Moreno et al., 2015) to 2–3 years in adolescents (Tierney et al., 2013, 2015). While not directly measuring speech perception, a recent study reported that 6-weeks of piano training enhanced the blood-oxygen-level-dependent (BOLD) response to melodies in fronto-parietal cortex and the cerebellum, implicated in the learning of auditory-motor associations (Herholz, Coffey, Pantev, & Zatorre, 2016). Elsewhere, as few as 2 weeks of piano-based training can lead to an enlargement of the MEG-measured auditory mismatch negativity (Lappe, Herholz, Trainor, & Pantev, 2008; Lappe, Trainor, Herholz, & Pantev, 2011), and 2 weeks of training to discriminate microtonal melodies can modulate the BOLD response in auditory cortex (Zatorre, Delhommeau, & Zarate, 2012). Taken together, the above findings provide evidence that short-term musical training can lead to neural auditory processing modulations which may benefit speech perception.

In older adults, cognitive training may be effective in offsetting age-related decline in cognitive abilities (such as memory and attention), and functional and structural changes – neuroplasticity – in brain regions associated with the trained task may follow a programme of training (Belleville & Bherer, 2012; Belleville et al., 2011; Belleville, Mellah, De Boysson, Demonet, & Bier, 2014; Zendel, de Boysson, Mellah, Démonet, & Belleville, 2016). In the auditory domain, non-musical training targeting abilities such as memory for sounds (e.g., repeating strings of words and syllables) and sound discrimination (e.g., speech syllable discrimination) appears to improve both SPIN perception and neural tracking of speech syllables (Anderson, White-Schwoch, Parbery-Clark, & Kraus, 2013). Memorizing and learning to discriminate sounds is important for musical training, but we are not presently aware of any study which has directly examined whether a programme of short-term musical training can alter the cortical processing of speech, and transfer to improved SPIN abilities in older adults. If short-term musical training can indeed transfer to SPIN perception, it could form an enjoyable ameliorative to speech perception difficulties in older adults, potentially contributing to improvements in quality of life.

In the present study, we recruited older adults with no prior musical experience, and randomly assigned them to receive either 6 months of piano-based music training, or to an active or passive control condition. In all conditions, fMRI data was acquired during a SPIN task from participants at two time points, at least 6-months apart (in the musical training and active control groups one of these sessions preceded the training programme, and the second session followed training). We were principally interested in the effect of music training on the BOLD signal and/or behavioural performance. It was expected that participants in the music training group would have an improved ability to understand SPIN and this would be related to a modulation in BOLD signal in brain areas involved in speech perception, reflecting a cross-domain (music-to-speech) transfer. In particular, we expected to see response modulations in regions involved in speech-motor integration due to their increased recruitment during speech processing in increasing background noise (Du, Buchsbaum, Grady, & Alain, 2014), particularly in older adults (Du, Buchsbaum, Grady, & Alain, 2016; Wong et al., 2009), and musicians (Du & Zatorre, 2017).

2. Materials and methods

Participants were enrolled in a broader study involving EEG, structural MRI, and functional MRI collection. Some of the structural MRI results have been published (West et al., 2017), and results from other parts of the study have recently been accepted for publication (Diarra et al., 2019; Zendel, West, Belleville, & Peretz, 2019). The participant recruitment and training procedures described therein are

the same as those described here.

2.1. Participants

Participants were recruited into the study from the Centre de Recherche, Institut Universitaire de Gériatrie de Montréal participant pool. The study received ethical approval from the Comité conjoint d'évaluation scientifique – Regroupement Neuroimagerie/Québec (CES-RNQ). All participants were native French speakers, and all testing was conducted in French. Participants were pre-screened to ensure that they did not have any present or past major illness, did not meet criteria for Mild Cognitive Impairment (MCI) using the Montreal Cognitive Assessment (Nasreddine et al., 2005), were not taking any psychiatric medications or medication known to have an impact on cognition, were MRI compatible, were a non-video game player and a non-musician. Pure-tone thresholds were measured before training to evaluate hearing status. All participants had a pure-tone average (PTA) for speech frequencies (500, 1000 & 2000 Hz) below 40 dB HL (i.e., normal hearing or mild hearing loss) in their best ear. To be considered a non-musician, participants had to not currently play a musical instrument, and had no more than 3 years of formal music training in their life. Music lessons that were part of the normal education curriculum were not included. People who sang regularly in a formal setting (i.e. choral singing) were considered musicians and were excluded from participation. To be considered a non-video game player, participants had little to no experience with commercial video games (e.g., games played on a computer or game console) during their lifetime. Casual games such as computerized card or puzzle games were not considered to be video games.

2.2. Experimental design

All participants were randomized into one of three groups. Randomization was done by an independent research assistant, using a predefined randomization table prior to contacting participants to ensure that participants were blind to the existence of the other two groups. Randomization was stratified using a covariate-adaptive randomization procedure. Each factor was stratified into two categories. For the factor of age there was “younger” (55–64 yrs) and “older” (65–75 yrs); for the factor of education there was low (< 16 yrs) and high (> 16 yrs); and for the factor of gender there was female and male. Because participants were recruited from a database, age, education level, and gender of each participant were known before they were contacted and it was thus possible to stratify randomization on the basis of these three factors. This stratification led to eight possible stratification groups. Lists of participants were provided from the participant database to the research assistant. Based on the stratification variables, the participant was assigned to one of the eight stratification groups based on the demographics available from the participant database (e.g. female, younger, high education; male, older, low education; etc...). Each of these eight groups was assigned a random but balanced order to determine experimental group assignment. That is the first person contacted who was in the “female, younger, higher education” stratification group was invited to participate in the experimental Music Training group (MUS; see below for details). If this participant accepted she became participant 1 in the MUS group. If she refused, the next person contacted in the “female, younger, higher education” stratification group was invited to participate in the MUS group. This repeated until a person in this stratification group volunteered to participate in the MUS group. Next, people in the same stratification group were invited to participate in the No Contact Control group (CON; see below for details) until one person volunteered to participate. Finally, a person in this stratification group was invited to participate in the Video Game training group (VID; see below for details) until one person volunteered to participate. This procedure was repeated, except the order of recruitment in experimental group was randomized for each cycle of

three assessments. That is each three participants were recruited into one of the three groups (MUS, VID, CON), but the order in which they were recruited was random. The orders were also randomized across all the stratification groups. Accordingly, participants who choose not to participate were not included in the randomization matrix.

Forty-five right handed participants in total were recruited into the study. Using the stratified randomization procedure, 15 participants were assigned to the videogame training (VID) group, 15 participants were assigned to the music training (MUS) group and 15 participants were assigned to the passive control (CON) group. One participant in the MUS group could not undergo MRI scanning because metal was discovered in their body just before the MRI scan. This participant completed the rest of the study, excluding the MRI portion. During the study, 2 further participants withdrew from the MUS group, 2 withdrew from the CON, while 11 withdrew from the VID group. Of the eleven participants that withdrew from the study, five withdrew consent to use their data, and therefore we cannot comment on their demographics. Of the six that did not make this request five were females and one was male. To account for the higher attrition rate within the VID group, an additional four participants were assigned, however, the stratified randomization procedure was not used. Of the four additional participants recruited into the VID group, two were males and two were females. This resulted in a total of 8 participants completing the training within the VID group. The demographics of the participants within each group are presented in Table 1.

2.3. Training procedure

2.3.1. Piano training group (MUS)

Piano training was done at home using Synthesia software, and an 88-key M-Audio MIDI piano. Synthesia is a piece of software that uses a non-standard form of musical notation that can be understood within a few minutes. This was critical as learning to read traditional music notation can take a long time. Notes in Synthesia are presented as coloured bars that fall from the top of the computer screen. At the bottom of the screen is an image of a piano keyboard, and when a coloured bar hits a certain note, the participant plays that note. The length of the bar indicates how long to hold the note for. First, the research assistant installed and calibrated the piano to work on the participant's home computer. Next, they completed an introductory lesson that included introductory information about music, detailed instructions on how to use Synthesia, and directions on how to record their progress. Introductory music information included lessons about note names, how to place hands on the piano, and how to synchronize performance with the information on the screen and the metronome. The RA worked with each participant until they were comfortable navigating the software interface, were able to complete the first lesson, and were able to load new songs or lessons. This typically took around 30–60 min. A set of introductory lessons and beginner piano music was installed on each participant's computer. The sixty-five introductory lessons were created by Artur Gajewski for use with Synthesia (2011). These lessons begin by playing a single note in time with a metronome, on each hand. They progress through alternating notes, playing notes simultaneously with both hands, scales (e.g., major and minor in different keys), chords, and the lessons end with a few etudes. Participants were told to start with the lessons and once they were comfortable with the lessons, to try out some of the introductory songs. Participants were encouraged to move

Table 1
Participant demographics.

	Age (± S.D.)	Education (± S.D.)	Gender (% of females)
VID Group	69.3 (5.7)	15.2 (3.2)	50%
MUS Group	67.7 (4.3)	14.7 (2.3)	83.3%
CON Group	66.9 (3.9)	17.5 (2.3)	76.9%

at their own pace, but to try to master a given lesson or song before moving on. Sometimes participants would work on a lesson and song simultaneously. The goal was to keep participants as engaged as possible in the piano lessons. While the selection of songs was at the discretion of the participant, all participants completed the introductory lessons. At the end of the training period, success was assessed by the highest lesson number each participant could complete with an accuracy score above 80%. Accuracy was defined as the number of notes the participant hit correctly during the lesson. For the participants in the Music group, the highest lesson with 80% accuracy ranged from lesson 10 to 60 ($M = 30.01$, $SD = 16.21$).

2.3.2. Video game training group (VID)

Video game training was done at home using the Nintendo Wii console system equipped with a Wii Classic Controller. All participants in this group trained on Super Mario 64. Two participants completed all task within Super Mario 64 before the completion of the 6-month training period. In these cases, they continued to train on a very similar game, Super Mario Galaxy, until the end of the training period. Super Mario 64 and Super Mario Galaxy are three-dimensional platform games where the player is tasked with exploring a virtual environment to search for stars (tokens). When enough stars are collected through completing in-game goals, the player can then progress further into the game and will encounter new environments to explore.

After the participant completed the pre-tests, a research assistant installed the Nintendo Wii to the participant's home television. The research assistant then gave an initial orientation to the participant to teach them how to turn on the Nintendo Wii and access the Super Mario 64 game. This was followed by a custom in-game orientation which taught the participant to move the character around the virtual environment. At this point, some participants encountered certain challenges associated with maneuvering the character. Some had issues with understanding the game's mechanics. Further, Super Mario 64 has a very steep learning curve that was not originally designed to be played by someone with little to no video game or computer experience. For this reason, the research assistant returned to the participant's home for up to three additional supervised 2-hour training sessions to teach the participant how to properly maneuver the character and progress through the game. After this, participants were given a custom made instruction booklet which outlined how and where to collect all the stars for the first four levels. This allowed participants to learn the game's mechanics in further detail and practice the basic motor coordination that was required. After this point, participants had to search for and obtain the stars within each remaining level without any assistance from the research team. All participants were able to find stars on their own, and progress through the game. Participants collected between 57 and 90 stars ($M = 69.1$, $SD = 13.5$) during the six-month training period.

2.3.3. No-contact control group (CON)

The passive control group had no contact with the research team during the six-month period other than to complete the pre-training, mid-term and post-testing sessions.

Music and Video game training lasted six months. In all cases, participants kept a record of their daily training progress and were asked to complete a minimum of 30 min of training at least five days a week, although some completed more than this amount. When they were recruited, all participants were instructed that they were in an experimental group, but they were blind to the existence of the other two groups. All participants were told that they were expected to improve in performance: Participants in the MUS group were told that we were examining the impact of music training on hearing abilities, and that the visual and cognitive tasks were included to determine how specific the benefit was; participants in the VID group were told that we were examining the impact of video game training on visual processing, and that the hearing and cognitive tasks were included to determine

how specific the training was; and the CON group was told we were examining how test-retest effects impact hearing, visual and cognitive testing. At the end of the study, all participants were debriefed and made aware of the other two groups.

2.4. Stimuli

All stimuli were presented via MRI-compatible Sensimetrics S14 insert earphones (Sensimetrics, USA). Sentences were taken from the French Canadian version of the Hearing in Noise test (Vaillancourt et al., 2005), and were presented in isolation at 75 dB SPL (No noise), or mixed with multi-talker babble at a signal-to-noise ratio of either 20 dB ("Quiet" noise) or 5 dB ("Loud" noise). These two conditions were normalized to 75 dB SPL after being combined with the noise. In addition, baseline trials (i.e., no stimulus) were included for all 3 noise conditions. Sentences were on average 1,876 ms (SD: 32 ms).

2.5. MRI data acquisition and task

In each scanning session, 5 runs of fMRI data were collected from each participant. Sessions took place at least 6 months apart, with identical procedures. Participants were scanned on a Siemens TIM Trio 3 T MRI system (Siemens Medical Solutions, Erlangen, Germany), using the Siemens 12-channel receive-only head coil at L'Unité de Neuroimagerie Fonctionnelle (UNF) of the Centre de recherche de l'Institut universitaire de gériatrie de Montréal. A mirror-box was attached to the head-coil, allowing participants to view picture arrays which were projected onto a screen at the head-end of the scanner bore and participants were given a button box on which to log their task responses. Each run consisted of a 34-vol, T2*-weighted, echo-planar imaging sequence (TR = 10,000 ms; TA = 2300 ms; TE = 30 ms; Flip Angle = 90°; matrix size = 64 × 64; in-plane resolution = 3 × 3 mm; slice thickness = 3 mm with 25% gap; number of slices = 36; acquisition plane = transverse; acquisition order = descending). A sparse-sampling method was employed to allow presentation of stimuli without scanner noise. On each trial, the participant heard 3000 ms of multi-talker babble noise, with a sentence occurring during the middle 2000 ms portion of the noise. The EPI T2* pulse sequence occurred 5500 ms after the onset of the sentence to best capture the auditory BOLD response (Hall et al., 2000). After each sentence, a 4-image picture array (example in Fig. 1) was presented, and participants were instructed to select the picture which most closely matched the meaning of the sentence (or indicate that they did not hear a sentence). The response image only included 4 options, with the fourth option always being "I did not hear a sentence". Participants were informed about this option before the scan, and to save space on the screen in the scanner, this option was presented at the bottom of the screen, leaving three 'larger' sentence options. Option #4 was considered correct on any trial where there was no sentence, although only trials where a speech stimulus was presented were considered in the behavioural analysis. 33 trials were completed in each of the 5 runs of functional scanning in each session.

A T1-weighted anatomical scan (TR = 2530 ms; TE = 1.64 ms; Flip Angle = 7°; matrix size = 256 × 256; in-plane resolution = 1 × 1 mm; slice thickness = 1 mm with no slice gap) consisting of 176 sagittal slices was collected in each session (except for a single scanning session of 1 participant, where 192 slices were acquired to maximize coverage), along with a gradient-echo 'fieldmapping' sequence which generated images containing information about inhomogeneities in the scanner's magnetic field (TR = 402 ms; Short TE = 4.92 ms; Long TE = 7.38 ms; Flip Angle = 60°; matrix size = 64 × 64; in-plane resolution = 3 × 3 mm; slice thickness = 3 mm with 25% gap; number of slices = 38; acquisition plane = transverse). In each session, a functional resting-state scan was also completed, along with scouts and localizers to guide the positioning of the acquisition matrix. These were not analyzed in the present study.

2.6. Data exclusions

Due to technical error, a single run of fMRI data was excluded for three participants (two from the MUS group, and one from the CON group) in one of their two scanning sessions. In two cases, the log file containing information about stimulus timing and behavioural performance was not properly generated by the presentation software; in these cases, the fMRI data was available for the runs in question, and was included in data preprocessing, but could not be included in statistical analysis. In the remaining case, neither the fMRI data nor the log file was available for the affected run. In summary, this resulted in the inclusion of 9 runs of data for the three affected participants in the final statistical analyses, where 10 runs were included in all other participants.

2.7. MRI data analysis

fMRI data were processed with SPM12-v6906 (<http://www.fil.ion.ucl.ac.uk/spm/>); Wellcome Trust Centre for Neuroimaging) under MATLAB R2016b on a desktop machine running Windows 8.1 (64-bit). Pre-processing consisted of standard routines, and proceeded separately for each session of scanning (i.e., one processing stream for the initial scanning session, and one processing stream for the second session). Initially, scans were reoriented so that the image origin sat on the anterior-posterior commissure axis. Next, voxel displacement maps were created from the gradient-echo fieldmapping sequences described above. These maps were then used to unwarp the functional scans and to realign them using an affine transformation consisting of 6 parameters, with each functional scan ultimately being realigned into the space of the first image collected during the first run of the session. The realignment stage resulted in resliced functional images, and also a mean image which was used to co-register the session-specific anatomical scans to their respective functional series. The co-registered anatomical images were then segmented into grey and white tissue probability maps, where forward deformation fields were also created. These deformation fields were used to spatially normalize functional (resampled voxel resolution = 3 × 3 × 3 mm) and anatomical (resampled voxel resolution = 1 × 1 × 1 mm) data to the Montreal Neurological Institute (MNI) standard space. Finally, functional data were smoothed with a gaussian kernel at 8 mm FWHM.

2.8. Statistical analyses

First-level models were specified in SPM12, and were run separately for each of the two imaging sessions. Participant-specific design matrices consisted of regressors for the No-noise, Quiet Noise and Loud Noise speech conditions, and for the 3 baseline conditions. The design matrices also contained 6 regressors modelling the motion parameters generated during the realignment preprocessing step. Serial correlations in the time series were accounted for by using SPM's default autoregressive AR(1) model during parameter estimation, and the data were high-pass filtered with a cut-off of 128 s. Stimulus onsets (specified in scans) were modelled as boxcars with a Finite Impulse Response function (FIR: window length = 5; order = 1). Three contrasts of interest were specified for each session, where each of the speech conditions was compared against the silent baseline. These contrasts were replicated across scanning runs, and scaled by the total number of runs collected for each participant to account for variability in the number of included scanning runs (see section on Data Exclusions above). In this manner, contrast vectors were multiplied by a value of $[1/n_{TotalRuns}]$ before being applied to estimated regression (beta) weights. An analysis of the unscaled fMRI data gave very similar results to the scaled data.

Whole-brain, group-level analysis was performed using the Sandwich Estimator (SwE) Toolbox (<http://www2.warwick.ac.uk/fac/sci/statistics/staff/academic-research/nichols/software/swe/>); Guillaume, Hua, Thompson, Waldorp, & Nichols, 2014). First-level

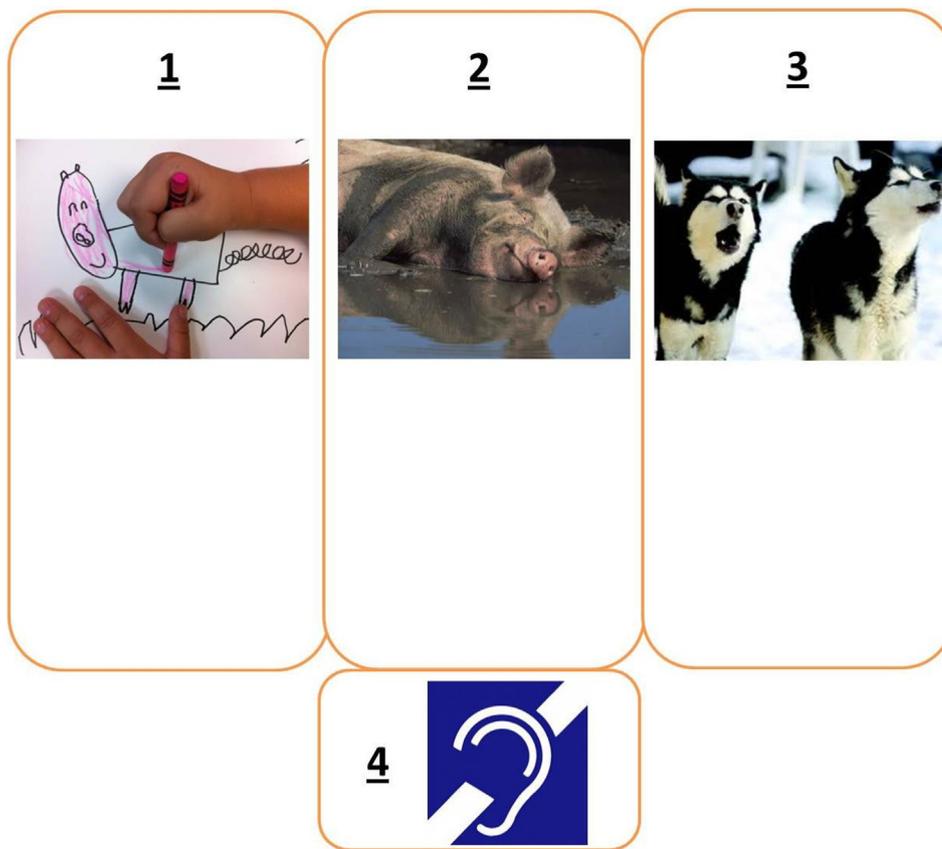


Fig. 1. Example of array presented during a typical trial during fMRI scanning. In this example, the participant heard “The child draws a pig” (in French), and then had to choose the picture which best matched the sentence. Participants were instructed to choose response option 4 when they could not hear a sentence, such as during a baseline trial.

contrast images were passed forward to a second-level factorial model, which tested for effects of Group, Session, and Condition, and all associated two- and three-way interactions. The model was estimated within a grey-matter mask which was derived from the grey-matter tissue probability map contained in SPM12 (thresholded at a probability of 0.2 and binarized). All second-level contrasts were corrected for multiple comparisons at the voxel level with a False Discovery Rate (FDR) threshold of $q = 0.05$.

We subsequently performed exploratory region-of-interest (ROI) analyses in 34 brain areas. While our hypotheses concerned behavioural improvement and modulations of BOLD activity connected to training, we initially selected a broad series of left-hemisphere regions implicated in three aspects of language processing (phonology, semantics, and sentence processing), as identified in previous meta-analyses (Vigneau et al., 2006, 2011). The lateralized selection was predicated on the stronger involvement of left-hemisphere auditory-motor regions in both melody processing in new piano learners (Herholz et al., 2016) and speech-in-noise processing in older listeners (Du et al., 2016), along with the apparent stronger left-lateralization for speech-motor processing in general (Hickok & Poeppel, 2007). Two of the ROI coordinate sets for the left Superior Temporal Gyrus specified in the meta-analyses were very proximal (MNI coordinates for phonological set: $-56, -12, -3$; semantic set: $-56, -13, -5$), and led to identical indices when transformed into the matrix space of the collected images. We therefore report only 29 left hemisphere ROI results, rather than the full set of 30, as listed in the meta-analyses (2011; Vigneau et al., 2006). However, we later expanded this selection to include 5 additional right-hemisphere regions which were recently shown to be more active in older adults as compared to younger adults, when both performed a phoneme-in-noise identification task with comparable accuracy (Du et al., 2016). Note that these right hemisphere co-ordinates were originally specified in Talairach space. These were transformed to approximate MNI coordinates using the MATLAB function ‘tal2mni’

(<http://imaging.mrcceb.cam.ac.uk/imaging/MniTalairach>). A complete list of the 34 included ROIs is presented in Table 2.

For each region, we created a sphere with a radius of 3 voxels around the reported peak coordinate, and extracted the mean contrast weight for the conditions of interest in each scanning session. The contrast weights were then analyzed in a 3 (Group) \times 2 (Session) \times 3 (Noise Level) mixed ANOVA with the ‘ezANOVA’ package in R (version 3.3.2). The behavioural accuracy data (averaged across runs for pre- and post-training testing sessions) were analyzed using the same ANOVA model specification described above. The Greenhouse-Geisser adjustment to the degrees of freedom was applied to those effects which failed Mauchly’s test of sphericity (i.e., p -value for the effect < 0.05). All obtained ROI p -values (for all 8 ANOVA effects: the main effects of group, session, and noise, the associated 2- and 3-way interactions, and the intercept/general effect term) were imported into MATLAB and a FDR threshold was calculated for multiple-comparisons correction using the ‘mafdr’ function implementing the procedure described by Benjamini and Hochberg (1995). We include the intercept/general effect term as it is appropriate to account for all effects tested by the model, but we briefly comment on the impact of removing this term upon the threshold in the results section. As we were principally interested in the effects of the type of training participants received, we focus our subsequent discussion on interaction effects involving both the training group and session factors.

3. Results

3.1. Behavioural results

Analysis of the behavioural accuracy on speech trials revealed a main effect of Noise Level ($F[2, 60] = 5.02, p = 0.01, \eta^2_{\text{partial}} = 0.14$), where performance was significantly better on the “Quiet” noise level (mean proportion correct collapsed across sessions = $0.84, \pm 0.02$ SE)

Table 2
Regions of Interest (ROIs) included in analyses (L/RH = Left/Right Hemisphere).

Source	ROI Label	MNI coordinate	
Vigneau et al. (2006) (LH: Phonological)	Rolandic Sulcus	-47, -6, 44	
	Precentral Gyrus	-48, 2, 26	
	Dorsal Pars Triangularis	-44, 23, 15	
	Rolandic Operculum	-48, 8, 3	
	Pars Orbitalis/Middle Frontal Gyrus	-33, 37, -6	
	Supramarginal Gyrus	-42, -52, 37	
	Superior Temporal Gyrus (STG)	-50, -38, 12	
	Planum Temporale	-60, -27, 9	
	Posterior Inferior Temporal Gyrus (ITG)	-50, -60, -7	
	Mid. Middle Temporal Gyrus (MTG)	-51, -35, -11	
Vigneau et al. (2006) (LH: Semantic)	Precentral Gyrus/P. Opercularis	-42, 4, 36	
	Dorsal Pars Opercularis	-44, 21, 24	
	Ventral Pars Triangularis	-43, 20, 4	
	Pars Orbitalis	-37, 31, -9	
	Angular Gyrus	-45, -68, 26	
	Posterior STG	-55, -48, 15	
	Anterior STG	-56, -13, -5	
	Posterior ITG	-46, -55, -7	
	Mid-lateral MTG	-59, -37, 1	
	Anterior Fusiform Gyrus	-38, -35, -13	
	Temporal Pole	-41, 3, -24	
	Vigneau et al. (2006) (LH: Sentence)	Posterior Middle Frontal Gyrus	-37, 10, 48
		Dorsal Pars Opercularis	-49, 16, 24
Ventral Pars Triangularis		-44, 26, 2	
Posterior Superior Temporal Sulcus		-50, -54, 22	
Anterior STG		-57, -13, -8	
Posterior MTG		-40, -63, 5	
Mid-lateral MTG		-57, -40, 2	
Temporal Pole		-47, 6, -24	
Du et al., 2016 (RH)	Pre/postcentral gyrus	33, -24, 48	
	STG/MTG	51, -14, -6	
	Medial Frontal Gyrus	12, -21, 53	
	Superior Parietal Lobule	18, -67, 57	
	Precuneus	14, -59, 21	

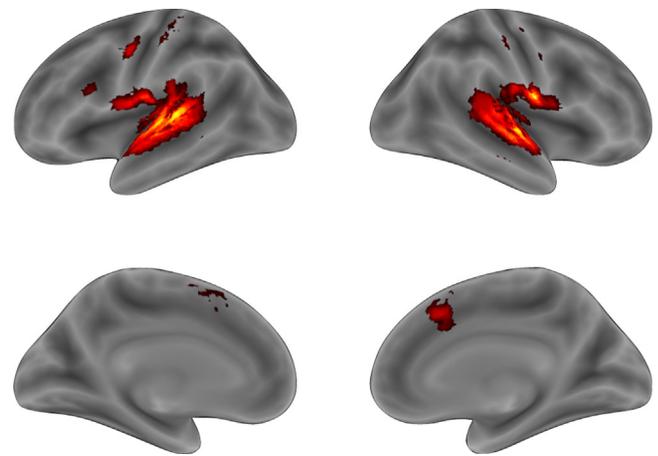


Fig. 3. Surface rendering showing areas where a significant main effect of noise level (at an FDR-corrected height threshold of 0.05, with arbitrary cluster threshold = 30 voxels) was found in the whole-brain analysis. Colour gradient (red to yellow) denotes the strength of the effect (X-statistic). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

enhanced as the level of babble noise increased (omnibus Main Effect of Noise Level derived from Group × Session × Noise Level factorial analysis: Loud > Quiet > No Noise; Note that the SwE toolbox uses the X-statistic for omnibus tests, in place of traditional F-statistics, and z-statistics in place of t-statistics for pairwise contrasts). The strongest responses which survived correction for multiple comparisons were found bilaterally in large clusters which included primary and secondary auditory areas (Heschl’s Gyrus, and the adjacent Planum Temporale and Planum Polare), the Superior Temporal Gyrus, and Opercular cortex (see Fig. 3 and Table 3 for a summary of activated peaks). As opposed to pairwise contrasts weighted towards increasing noise level, those weighted towards decreasing noise level (i.e., No Noise > Quiet Noise; No Noise > Loud Noise; Quiet Noise > Loud Noise) revealed no supra-threshold activation. Furthermore, the analysis revealed no supra-threshold effects of group or session, and no significant interactions between any of the factors.

3.3. ROI analysis

Exploratory analysis in ROIs derived from a meta-analysis of regions involved in speech processing (Vigneau et al., 2006) revealed an interaction of Group by Session in the left Middle Frontal Gyrus (MFG) ROI which survived the FDR correction procedure ($F[2, 30] = 5.79$, $p_{FDR} = 0.04$, $\eta^2_{\text{partial}} = 0.28$). In addition, we found weaker (based on effect size) Group × Session interactions which did not survive correction in left Rolandic Sulcus ($F[2, 30] = 3.88$, uncorrected $p = 0.03$, $\eta^2_{\text{partial}} = 0.21$), left Supramarginal Gyrus ($F[2, 30] = 4.39$, uncorrected $p = 0.02$, $\eta^2_{\text{partial}} = 0.23$), right Medial Frontal Gyrus ($F[2, 30] = 4.85$, uncorrected $p = 0.02$, $\eta^2_{\text{partial}} = 0.24$), and right STG/MTG ($F[2, 30] = 3.73$, uncorrected $p = 0.04$, $\eta^2_{\text{partial}} = 0.20$). The interaction effects did not appear to be contingent upon noise level, as indicated by the absence of significant three-way interaction effects in these regions. Plots of these results (Figs. 4 and 5) show that these interaction effects were driven by a greater post-training response increase in the MUS group as compared to the other two groups. Puzzlingly, we also noted apparent pre-training response differences between the groups (particularly in the frontal regions, between the MUS and CON groups), which contributed to the observed interaction effects. Importantly, however, the average response change in the MUS group always exceeded that of the other two groups in each ROI showing the group × session interaction.

We also found interactions of Session by Noise in left temporal

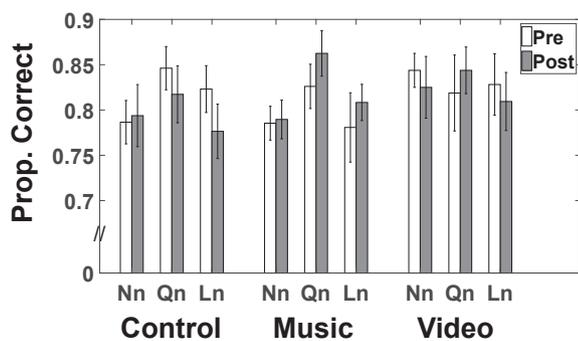


Fig. 2. Behavioural scores (overall proportion correct) on the in-scanner speech perception task. A significant main-effect of noise level was found, where participants performed better overall on in the Quiet noise condition. X-axis abbreviations: Nn = No Noise (i.e., speech with no background noise); Qn = ‘Quiet’ Noise (SNR = 20); Ln = ‘Loud’ Noise (SNR = 5). Error bars represent the standard error of the mean.

as compared to the “None” (mean = 0.80, ± 0.01 SE) and “Loud” (mean = 0.80, ± 0.01 SE) levels (Quiet > None, $t [32] = 3.72$, $p = 0.0008$; Quiet > Loud, $t [32] = 2.63$, $p = 0.01$). We found no significant interaction effects, nor any main effects of group or session. Behavioural results are presented in Fig. 2.

3.2. Whole-brain analysis

The BOLD response in bilateral temporal lobes responses was

Table 3

Whole-brain results Surviving voxel-level FDR correction (< 0.05 with arbitrary cluster threshold of 30 voxels). Key to abbreviations: Cent. Operc. = Central Operculum; PT = Planum Temporale; Precent. = Precentral; Paracing. = Paracingulate; IFG = Inferior Frontal Gyrus; Postcent. = Postcentral; MFG = Middle Frontal Gyrus; STG = Superior Temporal Gyrus. Note that for omnibus tests of main effects, the SWE toolbox gives the X-statistic rather than traditional F-statistics. Likewise, for pairwise contrasts, z-statistics are given in place of t-statistics.

Effect	MNI coordinates of cluster peak	Anatomical label of peak	Statistic	Cluster Size (#Voxels)	FDR-corrected p-value
Main Effect: Noise	[57, -10, 8]	Cent. Operc.	X = 74.96	1032	< 0.001
	[-60, -31, 8]	PT	X = 64.53	1044	< 0.001
	[-51, -7, 47]	Precent. Gyrus	X = 30.67	90	< 0.001
	[12, 17, 41]	Paracing. Gyrus	X = 23.49	147	< 0.001
	[54, -4, 50]	Precent. Gyrus	X = 20.35	50	< 0.001
	[-42, 11, 26]	IFG (operc.)	X = 16.21	32	0.001
Pairwise Contrast: Quiet Noise > None	[60, -13, 5]	PT	Z = 8.80	973	< 0.001
	[-51, -19, 8]	Heschl's Gyrus	Z = 7.43	1049	< 0.001
	[39, -58, -25]	Cerebellum	Z = 3.77	34	0.002
	[-39, -28, 62]	Postcent. Gyrus	Z = 3.77	36	0.002
	[51, -13, 53]	Postcent. Gyrus	Z = 3.73	44	0.002
Pairwise Contrast: Loud Noise > None	[57, -10, 8]	Cent. Operc.	Z = 9.00	1302	< 0.001
	[-60, -31, 8]	PT	Z = 8.22	1295	< 0.001
	[-51, -7, 47]	Precent. Gyrus	Z = 5.94	453	< 0.001
	[9, 20, 41]	Paracing. Gyrus	Z = 5.06	552	< 0.001
	[54, -4, 50]	Precent. Gyrus	Z = 4.54	258	< 0.001
	[24, -61, -22]	Cerebellum	Z = 3.44	38	0.005
	[42, 17, 26]	MFG	Z = 3.39	88	0.006
Pairwise Contrast: Loud Noise > Quiet Noise	[-66, -31, 8]	Post. STG	Z = 5.26	307	0.002
	[63, -25, 8]	Post. STG	Z = 4.86	434	0.002
	[9, 20, 41]	Paracing. Gyrus	Z = 4.60	62	0.002
	[-3, 11, 53]	Paracing. Gyrus	Z = 3.93	35	0.006

regions (STG, planum temporale – PT), and precentral gyrus; and main effects of Noise (i.e., reflecting the background-noise related increase observed in the whole-brain analysis) in temporal regions (left and right STG, MTG, and left PT), left rolandic regions (rolandic sulcus and operculum), and the pars opercularis of the left inferior frontal gyrus which survived correction (Table 4). We found no significant main effects of group or session, no significant interactions of Group by Noise, and no significant three-way interactions of the factors in any region, even at uncorrected thresholds. Note that excluding the intercept/general effect p-values from the FDR threshold calculation resulted in a

corrected p-value of 0.1 for the Group × Session interaction effect in left MFG. Nonetheless, as it seems appropriate to account for all effects tested by the model, we summarize all those effects which survived correction based on the threshold calculation which included the p-values for the intercept/general effect in Table 4.

3.4. Post-hoc brain-behaviour correlations

Following the initial analyses, we wished to determine whether performance changes in the behavioural task from pre-post session

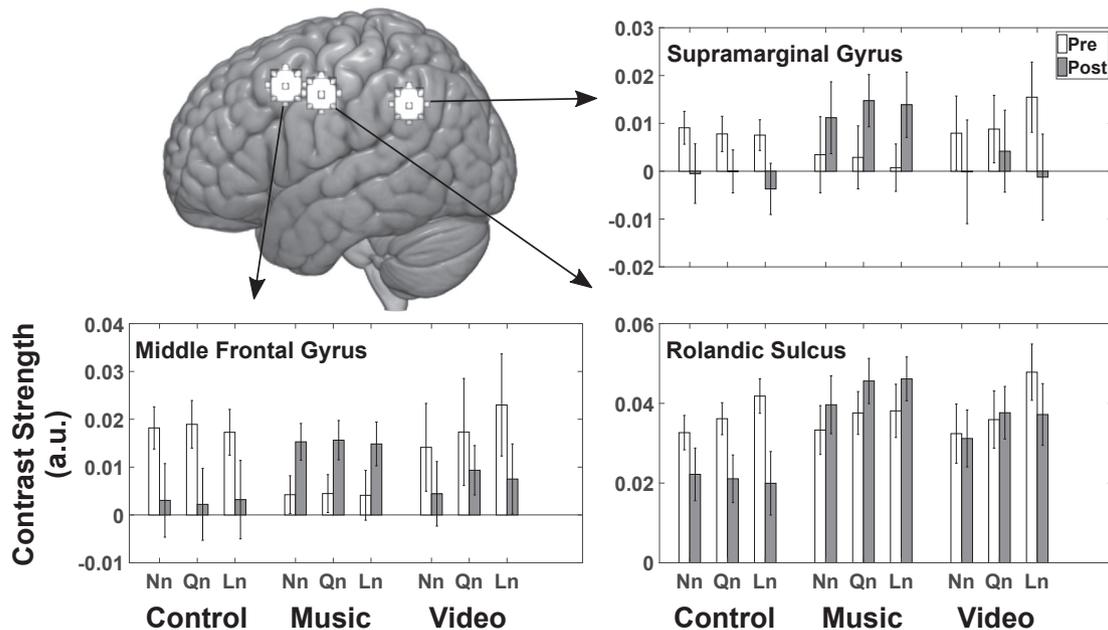


Fig. 4. Left hemisphere ROIs (superimposed on translucent brain template in MNI-space) showing Group × Session interactions as indicated by increased contrast weights post-training (grey bars) in the music group relative to the other two groups (bar plots showing scaled average contrast strength extracted from ROI spheres in arbitrary contrast units). The MFG interaction survived FDR-correction for multiple comparisons; the remaining p-values for the other two ROIs were < 0.05 at an uncorrected threshold. X-axis abbreviations are as in Fig. 2. Error bars represent the standard error of the mean.

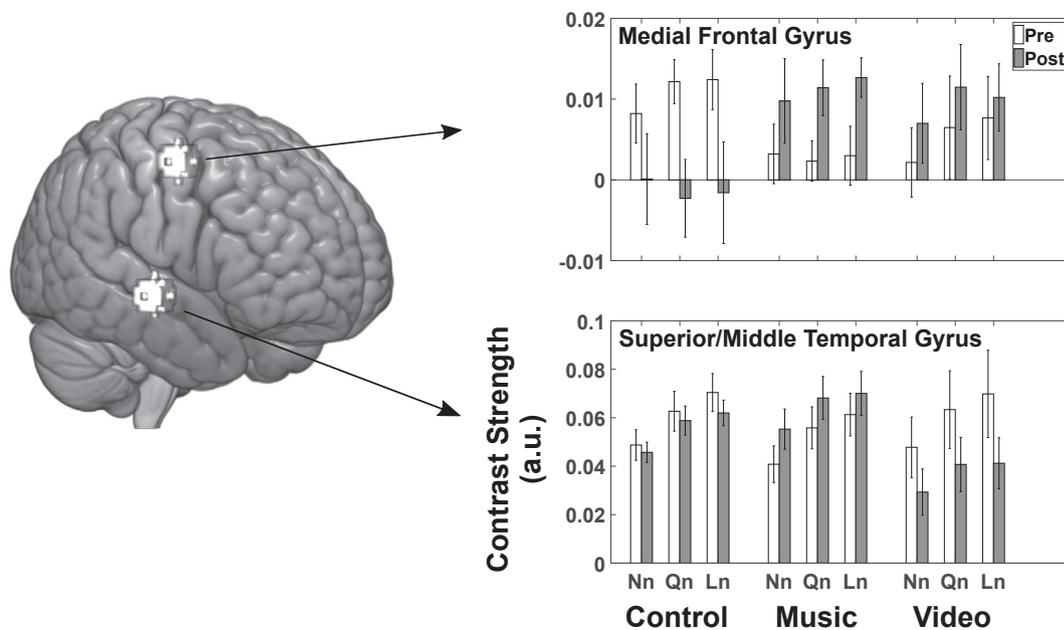


Fig. 5. Right hemisphere ROIs (superimposed on translucent brain template in MNI-space) showing Group \times Session interactions as indicated by increased contrast weights post-training (grey bars) in the MUS group relative to the other two groups (bar plots showing scaled average contrast strength extracted from ROI spheres arbitrary contrast units). P-values for these interactions were < 0.05 (uncorrected). X-axis abbreviations are as in Figs. 2 and 4. Error bars represent the standard error of the mean.

Table 4

ROI results. Key to abbreviations (only those not explained in Table 2): MTG = Middle Temporal Gyrus; Pars. Operc. = Pars Opercularis (of the IFG).

Effect	ROI	Hemisphere	MNI coordinates of sphere centre	F-Statistic	FDR-corrected p-value	Effect Size (partial eta-squared)
Interaction: Group \times Session	MFG	L	[-37, 10, 48]	5.79	0.04	0.28
Interaction: Session \times Noise	PT	L	[-60, -27, 9]	7.93	0.006	0.21
	Precent. Gyrus	L	[-48, 2, 26]	6.65	0.02	0.18
	STG	L	[-50, -38, 12]	8.77	0.004	0.23
	Ant. STG	L	[-56, -13, -5]	6.83	0.01	0.19
	Post. STG	L	[-55, -48, 15]	6.20	0.02	0.17
	STG	L	[-57, -13, -8]	5.62	0.03	0.16
Main Effect: Noise	PT	L	[-60, -27, 9]	167.11	< 0.0001	0.85
	Rolandic Operc.	L	[-48, 8, 3]	5.34	0.04	0.15
	Rolandic Sulcus	L	[-47, -6, 44]	9.80	0.002	0.25
	STG	L	[-50, -38, 12]	73.16	< 0.0001	0.71
	Mid-MTG	L	[-59, -37, 1]	52.77	< 0.0001	0.64
	Pars Operc.	L	[-44, 21, 24]	6.24	0.02	0.17
	Ant. STG	L	[-56, -13, -5]	113.02	< 0.0001	0.79
	Post. STG	L	[-55, -48, 15]	5.25	0.04	0.15
	Mid. MTG	L	[-57, -40, 2]	28.27	< 0.0001	0.49
	Pars Operc	L	[-49, 16, 24]	8.17	0.005	0.21
	STG	L	[-57, -13, -8]	75.60	< 0.0001	0.72
	STG/MTG	R	[50, -14, -4]	45.48	< 0.0001	0.60

correlated with the pre-post response changes in the ROIs described above. We therefore examined the relationship between behavioural accuracy and contrast weights within each ROI. We computed the difference between the mean post-training BOLD contrast strength and the mean pre-training contrast strength. We averaged across all noise levels as we found no significant three-way interaction in any of the tested ROIs, suggesting that the Group by Session interaction of interest was not significantly affected by variations in background noise level. This average difference value was then compared to the similarly extracted difference between mean ‘post’ and ‘pre’ behavioural accuracy (again averaged across noise levels) by estimating the Spearman correlation between the two. This analysis revealed significant positive correlations which survived FDR correction (FDR thresholds were calculated separately for each group’s vector of correlations) only in the MUS group, in the left MFG, and SMG (both rho coefficients = 0.85; FDR-corrected p-

values = 0.01; uncorrected p-values = 0.0008). The VID group also showed an MFG response-behaviour correlation, but this correlation did not survive FDR correction (rho = 0.83; corrected p-value = 0.17; uncorrected p-value = 0.02), unlike that found in the MUS group. The correlations are presented in Fig. 6.

3.5. Results summary

In the present study, we found that 6-months of musical training appeared to increase the response to speech in bilateral frontal (left Middle Frontal Gyrus and right Medial Frontal Gyrus), left parietal (left Supramarginal Gyrus), and right temporal (Superior/Middle Temporal Gyrus) cortical regions. Crucially, these enhancements were absent in both a passive control group, and an active control group who received 6 months of video-game training. While we did not find a clear

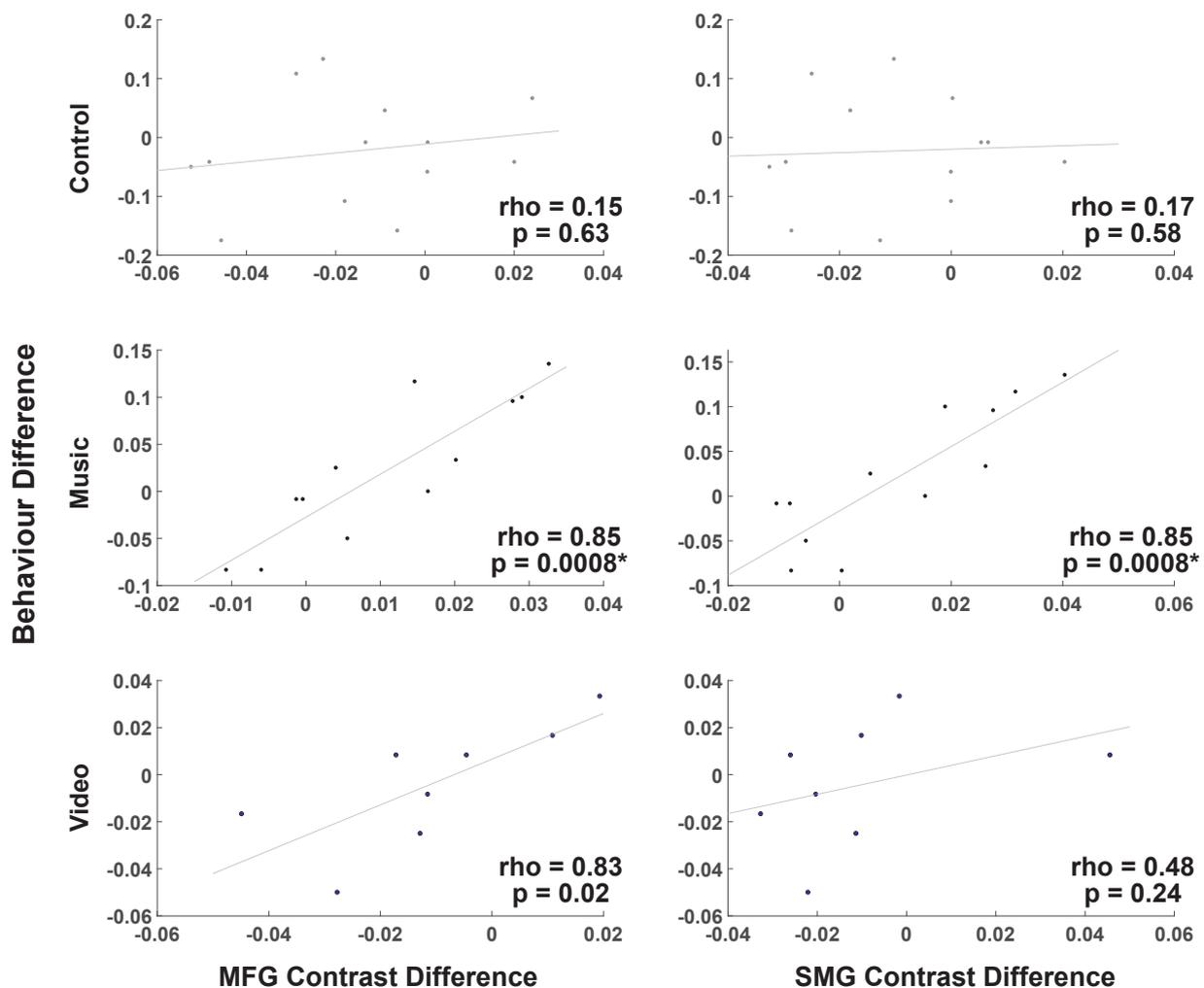


Fig. 6. Brain-behaviour correlations. For each session, behavioural performance and ROI contrast weights were averaged across noise levels, as we found no 3-way ROI interactions. The pre-session data was then subtracted from the post-session data, and the brain and behavioural data were correlated in all analyzed ROIs. In the MUS group (black dots), asterisks next to p-values denote significant brain-behaviour correlations which survived FDR correction and were found in the MFG (left panel) and SMG (right panel) ROIs. A similar correlation was found in the MFG for the VID group (blue dots), but this did not survive FDR correction. No significant correlations were found for the CON group (grey dots). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

improvement in behavioural performance, we did find that the response change from pre- to post-training in both the left Middle Frontal Gyrus (MFG) and the left Supramarginal Gyrus (SMG) was correlated with the pre- to post- change in training performance on the in-scanner speech perception task, and that these correlations survived correction only in the group which received musical training. Taken together with results obtained from the same participants during an EEG task (Zendel et al., 2019), these findings suggest that 6 months of musical training could transfer to speech perception, by modulating the neural response to speech in a way which might be beneficial for perception under noisy circumstances.

Our results are not without caveats: we have noted above the fact that there appear to be pre-training differences between the groups in the neural response to speech which contributed to the observed interaction effects in some measured regions, and that only the interaction effect observed in the MFG ROI survived correction for multiple comparisons. However, it is important to note that only the MUS group appeared to show a consistent post-training neural response enhancement across the regions showing these interaction effects. In light of this, and the fact that results from our correlation analyses are consistent with our hypothesis of post-training modulations which may benefit behaviour, we offer cautious interpretation in the following

sections. Our discussion will focus mainly on the effects observed in the left MFG and SMG regions, as these regions show at least either a significant interaction effect, or a significant brain-behaviour correlation surviving correction.

4. Discussion

4.1. Training-related cortical response changes which track SPIN performance changes in left fronto-parietal regions

Music training increased the response to speech in part of the posterior left MFG (BA 6, 8, 9, and 46), an area overlapping with dorso-lateral prefrontal cortex (dlPFC) regions involved in working memory and top-down control (Bengtsson et al., 2009; Constantinidis & Klingberg, 2016; Dosenbach et al., 2006; Dosenbach, Fair, Cohen, Schlaggar, & Petersen, 2008; Ester, Sprague, & Serences, 2015; Giordano et al., 2014; Kim, Kroger, Calhoun, & Clark, 2015; Vogan, Morgan, Powell, Smith, & Taylor, 2016). This area appears to be involved in monitoring errors on short timescales (Dosenbach et al., 2008), which would explain the observed training-related functional plasticity: assessing self-generated piano sequences for errors led to response increases in control regions (Pallesen et al., 2010) which

transferred to a task involving careful monitoring of speech. Previous training studies with older participants have revealed plasticity in PFC regions including the MFG (Belleville et al., 2011, 2014), supporting the notion of the area as a generic centre for cognitive control (Dosenbach et al., 2008, 2006). A general training effect could therefore explain these results, especially given that the relationship between SPIN performance change and contrast change was found in both active training groups, although we note that the overall direction of brain contrast changes differed between the MUS and VID groups (response increases in the MUS group, and decreases in the VID group), and the brain-behaviour correlation only survived correction for multiple comparisons in the MUS group. As this appears to be attributable to the low sample size in this condition, we cannot dismiss the possibility (introduced above) that effects in the MFG are not entirely explained by music training *per se*, but by general training involving the control and monitoring of perceptual-motor processes (be they auditory or visual). The similarity between the MUS and VID group correlations was not, however, reflected in the second ROI where a brain-behaviour relationship was found.

In the parietal lobe, the SMG's output during musical perception and motor execution is modulated by musical training in both the longer (Bangert et al., 2006; Bengtsson & Ullén, 2006; Kleber, Friberg, Zeitouni, & Zatorre, 2017) and shorter term (Herholz et al., 2016), and may serve as a hub for processing sung or spoken sensorimotor feedback information from the vocal apparatus (Golfopoulos et al., 2011; Kleber et al., 2017; Tourville, Reilly, & Guenther, 2008). The training in the present study likely led to the formation of new sensorimotor relationships; i.e., an increased ability to relate motor performance (playing a note) to auditory percept (hearing a note). This may have sharpened the SMG's role in attending to, and discriminating speech sounds (Hartwigsen et al., 2010; Vigneau et al., 2006), perhaps through a covert "sounding out" of the presented speech. This is also reflected in the observed post-training response increases in the Rolandic Sulcus, which overlapped closely with regions (BAs 4 and 6) involved in both the perception and production of speech (Du et al., 2014; Wilson, Saygin, Sereno, & Iacoboni, 2004). This region is close to precentral gyrus areas whose response during musical perception and motor performance is enhanced in professional piano players as compared with non-musicians (Bangert et al., 2006), further suggestive of training-related increases in speech-motor regions which may transfer to speech perception. Together, these three left-hemisphere regions have previously been delineated as part of a fronto-parietal language network (Catani, Jones, & Ffytche, 2005; López-Barroso et al., 2015; Vigneau et al., 2006). Correlations between improved SPIN understanding and enhanced activity in MFG and SMG provide further support that these regions are involved in speech perception, although we interpret these results with caution given the post-hoc and exploratory nature of the analysis.

Lastly, we also observed interaction effects in right hemisphere regions which were consistent with those found in the abovementioned left hemisphere ROIs. However, none of these regions showed either (A) an interaction effect which survived FDR correction; nor (B) a significant correlation with behaviour which survived correction. We therefore make no strong claims about the influence of these regions in the present study, beyond stating that the weak evidence of their involvement is consistent with previous results: The right Medial Frontal Gyrus ROI (BAs 4 and 6) is located closely to parts of the Supplementary Motor Area (SMA), which has variously been implicated in the perception and production of speech (Kotz & Schwartz, 2010) and in verbal and tonal working memory in musicians (Schulze, Zysset, Mueller, Friederici, & Koelsch, 2011). Furthermore, right hemisphere-derived auditory ERPs are enhanced in lifelong musicians during sound discrimination (Zendel & Alain, 2014), and short-term functional plasticity in right hemisphere auditory cortex has been shown following training in pitch discrimination (Ley et al., 2012; Zatorre et al., 2012). These findings provide some context for the involvement of right

Superior/Middle Temporal Gyrus (BA 21/22 in the ROI) current study.

4.2. Could the observed effects be driven by differences in training difficulty?

So far, we have not discussed the issue of differences in training difficulty between music and videogame programmes: based on limited participant feedback, it appears that the VID participants found their training quite challenging. However, it should be noted that the direction of neural results for the CON & VID groups is similar in the MFG, STG/MTG, and SMG, where the pattern for both is distinct from that present in the MUS group. While it is possible that training difficulty may be driving these effects we would expect differences in brain activation between the CON (where there was no training at all) and VID groups if this were the case.

Meanwhile, brain results for the MUS & VID groups appear similar in Med. FG and Rolandic Sulcus regions (both differing from the CON group). These areas are more directly implicated in motor processes and might show general enhancements for any task involving sharpening of such processes (mapping piano keypresses vs. mapping videogame keypad presses). The similarities between the MUS and VID groups in these areas do not suggest that perceived difficulty is influencing the pattern of activation, but this is harder to interpret for MFG, which is supposedly domain-general (Dosenbach et al., 2008, 2006). However, again, if training difficulty played a strong role, then we would expect differences between not only MUS and VID groups, but also CON and VID groups. On the other hand, we might expect to see different patterns between the MUS and VID groups in SMG and STG/MTG simply because there was more auditory engagement in the music training than the videogame training and not because of training difficulty *per se*.

4.3. The absence of a direct behavioural benefit

Despite the neural effects and their relationship to behaviour, we found no clear effect of training on behaviour, and, surprisingly, that accuracy was highest in the quieter background noise condition. The latter effect was likely due to experimental design: Noise levels were randomized within a block so that, on any given trial with background noise (Quiet & Loud), the noise started before the target speech, and therefore may have acted as a cue. In quieter noise, this did not interfere with understanding, while the cue provided by the noise signalled the upcoming trial. The no-noise condition did not have this cue. The long inter-stimulus onset interval (10 s) may have led to an attentional 'drifting' during the no-noise trials, possibly affecting the ability to understand the start of the sentence. The lack of training effect was also puzzling as the same MUS participants showed a significant training effect during the EEG task (Zendel et al., 2019). There are two possible reasons for the absence of an in-scanner training effect: firstly, higher SNRs were employed during the loud and quiet noise conditions (5 vs 0 dB SNR [loud] and 20 vs 15 dB SNR [quiet], for fMRI and EEG, respectively). In the EEG study a training benefit was only observed for the loudest noise (0 dB SNR) condition, but single words without context were used, while complete sentences were used in the present fMRI study. Secondly, the EEG study required participants to repeat the word aloud, as compared to the forced-choice task used here. The difference in pattern of results suggests that short-term music training may only benefit the ability to understand speech in the most difficult contexts, where participants are given no cues (such as pictures) to the correct response. Nevertheless, these findings suggest that music training enhances neural responses to speech in speech-motor and control regions. The absence of training-related interactions with noise level suggests that the benefit of music training is related to processing speech in general, and not separating it from background noise; however due to ceiling effects, the ability to understand SPIN is only observed in the most challenging listening situations. Previous work comparing older

musicians and non-musicians using behavioural techniques alone would not have been sensitive to this pattern because performance without noise is always near ceiling (Parbery-Clark et al., 2011; Zendel & Alain, 2012).

4.4. Conclusion

One of the key skills acquired during musical training is strengthening the connection between the auditory and motor systems, as a musician must learn to adjust motor movements based on incoming auditory information. This is a critical first step when learning music, incoming auditory information must be processed in conjunction with the error-monitoring system to detect mistakes, and then transmit this information to the motor system for corrective updates. Interpreted within this scheme, our results are consistent with explanatory frameworks such as the OPERA hypothesis (Patel, 2011, 2012), which links enhanced neural coding of speech with the demands placed on the auditory and motor systems during musical training and performance. While we do not find a clear behavioural improvement, the auditory-motor activity increases shown in our music training group do appear to track behavioural performance to some extent, and this same group showed improvement on a different speech perception task (Zendel et al., 2019). These observations suggest that musical training in old age may be a fruitful and enjoyable means of countering aspects of age-related decline in SPIN perception.

Acknowledgements

Support for this research came from the Canada Research Chairs Program (IP & BRZ), the GRAMMY Foundation (BRZ), Fondation Caroline Durant (SB), Fonds de recherche du Québec-Santé (FRQS) (BRZ), Natural Sciences and Engineering Research Council of Canada (SB), and Natural Sciences and Engineering Research Council of Canada Collaborative Research and Training Experience Program in Auditory Cognitive Neuroscience (NSERC-CREATE-ACN) (BRZ). The authors declare no competing financial interests.

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.bandc.2019.103592>.

References

- Anderson, S., White-Schwoch, T., Parbery-Clark, A., & Kraus, N. (2013). Reversal of age-related neural timing delays with training. *Proceedings of the National Academy of Sciences*, 110(11), 4357–4362. <https://doi.org/10.1073/pnas.1213555110>.
- Bangert, M., Peschel, T., Schlaug, G., Rotte, M., Drescher, D., Hinrichs, H., ... Altenmüller, E. (2006). Shared networks for auditory and motor processing in professional pianists: Evidence from fMRI conjunction. *NeuroImage*, 30(3), 917–926. <https://doi.org/10.1016/j.neuroimage.2005.10.044>.
- Belleville, S., & Bherer, L. (2012). Biomarkers of cognitive training effects in aging. *Current Translational Geriatrics and Experimental Gerontology Reports*, 1(2), 104–110. <https://doi.org/10.1007/s13670-012-0014-5>.
- Belleville, S., Clément, F., Mella, S., Gilbert, B., Fontaine, F., & Gauthier, S. (2011). Training-related brain plasticity in subjects at risk of developing Alzheimer's disease. *Brain*, 134(6), 1623–1634. <https://doi.org/10.1093/brain/awr037>.
- Belleville, S., Mella, S., De Boissson, C., Demonet, J. F., & Bier, B. (2014). The pattern and loci of training-induced brain changes in healthy older adults are predicted by the nature of the intervention. *PLoS ONE*, 9(8), <https://doi.org/10.1371/journal.pone.0102710>.
- Bengtsson, S. L., & Ullén, F. (2006). Dissociation between melodic and rhythmic processing during piano performance from musical scores. *NeuroImage*, 30(1), 272–284. <https://doi.org/10.1016/j.neuroimage.2005.09.019>.
- Bengtsson, S. L., Ullén, F., Henrik Ehrsson, H., Hashimoto, T., Kito, T., Naito, E., ... Sadato, N. (2009). Listening to rhythms activates motor and premotor cortices. *Cortex*, 45(1), 62–71. <https://doi.org/10.1016/j.cortex.2008.07.002>.
- Benjamini, Y., & Hochberg, Y. (1995). Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B*, 57(1), 289–300. <https://doi.org/10.2307/2346101>.
- Catani, M., Jones, D. K., & Fyfe, D. H. (2005). Perisylvian language networks of the human brain. *Annals of Neurology*, 57(1), 8–16. <https://doi.org/10.1002/ana.20319>.
- Coffey, E. B. J., Mogilever, N. B., & Zatorre, R. J. (2017). Speech-in-noise perception in musicians: A review. *Hearing Research*, 352, 49–69. <https://doi.org/10.1016/j.heares.2017.02.006>.
- Constantinidis, C., & Klingberg, T. (2016). The neuroscience of working memory capacity and training. *Nature Reviews Neuroscience*, 17(7), 438–449. <https://doi.org/10.1038/nrn.2016.43>.
- Diarra, M., Zendel, B. R., Benady-Chorney, J., Blanchette, C. A., Lepore, F., Peretz, I., ... West, G. L. (2019). Playing Super Mario increases oculomotor inhibition and frontal eye field grey matter in older adults. *Experimental Brain Research*, 237(3), 723–733. <https://doi.org/10.1007/s00221-018-5453-6>.
- Dosenbach, N. U. F., Fair, D. A., Cohen, A. L., Schlaggar, B. L., & Petersen, S. E. (2008). A dual-networks architecture of top-down control. *Trends in Cognitive Sciences*, 12(3), 99–105. <https://doi.org/10.1016/j.tics.2008.01.001>.
- Dosenbach, N. U. F., Visscher, K. M., Palmer, E. D., Miezin, F. M., Wenger, K. K., Kang, H. C., ... Petersen, S. E. (2006). A core system for the implementation of task sets. *Neuron*, 50(5), 799–812. <https://doi.org/10.1016/j.neuron.2006.04.031>.
- Du, Y., Buchsbaum, B. R., Grady, C. L., & Alain, C. (2014). Noise differentially impacts phoneme representations in the auditory and speech motor systems. *Proceedings of the National Academy of Sciences*, 111(19), 7126–7131. <https://doi.org/10.1073/pnas.1318738111>.
- Du, Y., Buchsbaum, B. R., Grady, C. L., & Alain, C. (2016). Increased activity in frontal motor cortex compensates impaired speech perception in older adults. *Nature Communications*, 7, 12241. <https://doi.org/10.1038/ncomms12241>.
- Du, Y., & Zatorre, R. J. (2017). Musical training sharpens and bonds ears and tongue to hear speech better. *Proceedings of the National Academy of Sciences*, 201712223. <https://doi.org/10.1073/pnas.1712223114>.
- Ester, E. F., Sprague, T. C., & Serences, J. T. (2015). Parietal and frontal cortex encode stimulus-specific mnemonic representations during visual working memory. *Neuron*, 87(4), 893–905. <https://doi.org/10.1016/j.neuron.2015.07.013>.
- Fujioka, T., Ross, B., Kakigi, R., Pantev, C., & Trainor, L. J. (2006). One year of musical training affects development of auditory cortical-evoked fields in young children. *Brain*, 129(10), 2593–2608. <https://doi.org/10.1093/brain/awl247>.
- Giordano, B. L., Pernet, C., Charest, I., Belizaire, G., Zatorre, R. J., & Belin, P. (2014). Automatic domain-general processing of sound source identity in the left posterior middle frontal gyrus. *Cortex*, 58, 170–185. <https://doi.org/10.1016/j.cortex.2014.06.005>.
- Golfinoopoulos, E., Tourville, J. A., Bohland, J. W., Ghosh, S. S., Nieto-Castanon, A., & Guenther, F. H. (2011). fMRI investigation of unexpected somatosensory feedback perturbation during speech. *NeuroImage*, 55(3), 1324–1338. <https://doi.org/10.1016/j.neuroimage.2010.12.065>.
- Guillaume, B., Hua, X., Thompson, P. M., Waldorp, L., & Nichols, T. E. (2014). Fast and accurate modelling of longitudinal and repeated measures neuroimaging data. *NeuroImage*, 94, 287–302. <https://doi.org/10.1016/j.neuroimage.2014.03.029>.
- Hall, D. A., Summerfield, A. Q., Gonçalves, M. S., Foster, J. R., Palmer, A. R., & Bowtell, R. W. (2000). Time-course of the auditory BOLD response to scanner noise. *Magnetic Resonance in Medicine*, 43(4), 601–606. [https://doi.org/10.1002/\(SICI\)1522-2594\(200004\)43:4<601::AID-MRM16>3.0.CO;2-R](https://doi.org/10.1002/(SICI)1522-2594(200004)43:4<601::AID-MRM16>3.0.CO;2-R).
- Hartwigsen, G., Baumgaertner, A., Price, C. J., Koehnke, M., Ulmer, S., & Siebner, H. R. (2010). Phonological decisions require both the left and right supramarginal gyri. *Proceedings of the National Academy of Sciences*, 107(38), 16494–16499. <https://doi.org/10.1073/pnas.1008121107>.
- Herholz, S. C., Coffey, E. B. J., Pantev, C., & Zatorre, R. J. (2016). Dissociation of neural networks for predisposition and for training-related plasticity in auditory-motor learning. *Cerebral Cortex*, 26(7), 3125–3134. <https://doi.org/10.1093/cercor/bhv138>.
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Neuroscience*, 8(May), 393–402.
- Kim, C., Kroger, J. K., Calhoun, V. D., & Clark, V. P. (2015). The role of the frontopolar cortex in manipulation of integrated information in working memory. *Neuroscience Letters*, 595, 25–29. <https://doi.org/10.1016/j.neulet.2015.03.044>.
- Kleber, B., Friberg, A., Zeitouni, A., & Zatorre, R. (2017). Experience-dependent modulation of right anterior insula and sensorimotor regions as a function of noise-masked auditory feedback in singers and nonsingers. *NeuroImage*, 147(November 2016), 97–110. <https://doi.org/10.1016/j.neuroimage.2016.11.059>.
- Kotz, S. A., & Schwartz, M. (2010). Cortical speech processing unplugged: A timely subcortico-cortical framework. *Trends in Cognitive Sciences*, 14(9), 392–399. <https://doi.org/10.1016/j.tics.2010.06.005>.
- Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience*, 11(8), 599–605.
- Lappe, C., Herholz, S. C., Trainor, L. J., & Pantev, C. (2008). Cortical plasticity induced by short-term unimodal and multimodal musical training. *Journal of Neuroscience*, 28(39), 9632–9639. <https://doi.org/10.1523/JNEUROSCI.2254-08.2008>.
- Lappe, C., Trainor, L. J., Herholz, S. C., & Pantev, C. (2011). Cortical plasticity induced by short-term multimodal musical rhythm training. *PLoS ONE*, 6(6), <https://doi.org/10.1371/journal.pone.0021493>.
- Ley, A., Vroomen, J., Hausfeld, L., Valente, G., De Weerd, P., & Formisano, E. (2012). Learning of new sound categories shapes neural response patterns in human auditory cortex. *Journal of Neuroscience*, 32(38), 13273–13280. <https://doi.org/10.1523/JNEUROSCI.0584-12.2012>.
- López-Barroso, D., Ripollés, P., Marco-Pallarés, J., Mohammadi, B., Münte, T. F., Bachoud-Lévi, A. C., ... de Diego-Balaguer, R. (2015). Multiple brain networks underpinning word learning from fluent speech revealed by independent component analysis. *NeuroImage*, 110, 182–193. <https://doi.org/10.1016/j.neuroimage.2014.12.085>.
- Moreno, S., & Besson, M. (2006). Musical training and language-related brain electrical activity in children. *Psychophysiology*, 43(3), 287–291. <https://doi.org/10.1111/j>

- 1469-8986.2006.00401.x.
- Moreno, S., Lee, Y., Janus, M., & Bialystok, E. (2015). Short-term second language and music training induces lasting functional brain changes in early childhood. *Child Development, 86*(2), 394–406. <https://doi.org/10.1111/cdev.12297>.
- Moreno, S., Marques, C., & Santos, A. (2009). Musical training influences linguistic abilities in 8-year-old children: More evidence for brain plasticity. *Cerebral Cortex, 19*(3), 712–723. <https://doi.org/10.1093/cercor/bhn120>.
- Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., ... Chertkow, H. (2005). The Montreal cognitive assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society, 53*(4), 695–699. <https://doi.org/10.1111/j.1532-5415.2005.53221.x>.
- Pallesen, K. J., Brattico, E., Bailey, C. J., Korvenoja, A., Koivisto, J., Gjedde, A., & Carlson, S. (2010). Cognitive control in auditory working memory is enhanced in musicians. *PLoS ONE, 5*(6), <https://doi.org/10.1371/journal.pone.0011120>.
- Parbery-Clark, A., Anderson, S., Hittner, E., & Kraus, N. (2012). Musical experience offsets age-related delays in neural timing. *Neurobiology of Aging, 33*(7), 1483.e1–1483.e4. <https://doi.org/10.1016/j.neurobiolaging.2011.12.015>.
- Parbery-Clark, A., Strait, D. L., Anderson, S., Hittner, E., & Kraus, N. (2011). Musical experience and the aging auditory system: Implications for cognitive abilities and hearing speech in noise. *PLoS ONE, 6*(5), <https://doi.org/10.1371/journal.pone.0018082>.
- Parbery-Clark, A., Strait, D. L., & Kraus, N. (2011). Context-dependent encoding in the auditory brainstem subserves enhanced speech-in-noise perception in musicians. *Neuropsychologia, 49*(12), 3338–3345. <https://doi.org/10.1016/j.neuropsychologia.2011.08.007>.
- Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Frontiers in Psychology, 2*(JUN), 1–14. <https://doi.org/10.3389/fpsyg.2011.00142>.
- Patel, A. D. (2012). The OPERA hypothesis: Assumptions and clarifications. *Annals of the New York Academy of Sciences, 1252*(1), 124–128. <https://doi.org/10.1111/j.1749-6632.2011.06426.x>.
- Schulze, K., Zysset, S., Mueller, K., Friederici, A. D., & Koelsch, S. (2011). Neuroarchitecture of verbal and tonal working memory in nonmusicians and musicians. *Human Brain Mapping, 32*(5), 771–783. <https://doi.org/10.1002/hbm.21060>.
- Slater, J., Skoe, E., Strait, D. L., O'Connell, S., Thompson, E., & Kraus, N. (2015). Music training improves speech-in-noise perception: Longitudinal evidence from a community-based music program. *Behavioural Brain Research, 291*, 244–252. <https://doi.org/10.1016/j.bbr.2015.05.026>.
- Tierney, A. T., Krizman, J., & Kraus, N. (2015). Music training alters the course of adolescent auditory development. *Proceedings of the National Academy of Sciences, 112*(32), 10062–10067. <https://doi.org/10.1073/pnas.1505114112>.
- Tierney, A., Krizman, J., Skoe, E., Johnston, K., & Kraus, N. (2013). High school music classes enhance the neural processing of speech. *Frontiers in Psychology, 4*(DEC), 1–7. <https://doi.org/10.3389/fpsyg.2013.00855>.
- Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *NeuroImage, 39*(3), 1429–1443. <https://doi.org/10.1016/j.neuroimage.2007.09.054>.
- Vaillancourt, V., Laroche, C., Mayer, C., Basque, C., Nali, M., Eriks-Brophy, A., ... Giguère, C. (2005). Adaptation of the HINT (hearing in noise test) for adult Canadian Francophone populations. *International Journal of Audiology, 44*(6), 358–369. <https://doi.org/10.1080/14992020500060875>.
- Vigneau, M., Beaucois, V., Hervé, P. Y., Duffau, H., Crivello, F., Houdé, O., ... Tzourio-Mazoyer, N. (2006). Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing. *NeuroImage, 30*(4), 1414–1432. <https://doi.org/10.1016/j.neuroimage.2005.11.002>.
- Vigneau, M., Beaucois, V., Hervé, P. Y., Jobard, G., Petit, L., Crivello, F., ... Tzourio-Mazoyer, N. (2011). What is right-hemisphere contribution to phonological, lexico-semantic, and sentence processing? Insights from a meta-analysis. *NeuroImage, 54*(1), 577–593. <https://doi.org/10.1016/j.neuroimage.2010.07.036>.
- Vogan, V. M., Morgan, B. R., Powell, T. L., Smith, M. L., & Taylor, M. J. (2016). The neurodevelopmental differences of increasing verbal working memory demand in children and adults. *Developmental Cognitive Neuroscience, 17*, 19–27. <https://doi.org/10.1016/j.dcn.2015.10.008>.
- West, G. L., Zengel, B. R., Konishi, K., Benady-Chorney, J., Bohbot, V. D., Peretz, I., & Belleville, S. (2017). Playing Super Mario 64 increases hippocampal grey matter in older adults. *PLoS ONE, 12*(12), 1–18. <https://doi.org/10.1371/journal.pone.0187779>.
- Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience, 7*(7), 701–702. <https://doi.org/10.1038/nn1263>.
- Wong, P. C. M., Jin, J. X., Gunasekera, G. M., Abel, R., Lee, E. R., & Dhar, S. (2009). Aging and cortical mechanisms of speech perception in noise. *Neuropsychologia, 47*(3), 693–703. <https://doi.org/10.1016/j.neuropsychologia.2008.11.032>.
- Zatorre, R. J., Delhommeau, K., & Zarate, J. M. (2012). Modulation of auditory cortex response to pitch variation following training with microtonal melodies. *Frontiers in Psychology, 3*(DEC), 1–17. <https://doi.org/10.3389/fpsyg.2012.00544>.
- Zengel, B. R., & Alain, C. (2012). Musicians experience less age-related decline in central auditory processing. *Psychology and Aging, 27*(2), 410–417. <https://doi.org/10.1037/a0024816>.
- Zengel, B. R., & Alain, C. (2014). Enhanced attention-dependent activity in the auditory cortex of older musicians. *Neurobiology of Aging, 35*(1), 55–63. <https://doi.org/10.1016/j.neurobiolaging.2013.06.022>.
- Zengel, B. R., de Boysson, C., Mellah, S., Démonet, J. F., & Belleville, S. (2016). The impact of attentional training on event-related potentials in older adults. *Neurobiology of Aging, 47*, 10–22. <https://doi.org/10.1016/j.neurobiolaging.2016.06.023>.
- Zengel, B. R., Tremblay, C.-D., Belleville, S., & Peretz, I. (2015). The impact of musicianship on the cortical mechanisms related to separating speech from background noise. *Journal of Cognitive Neuroscience, 27*(5), 1044–1059. https://doi.org/10.1162/jocn_a_00758.
- Zengel, B. R., West, G., Belleville, S., & Peretz, I. (2019). Music training improves the ability to understand speech-in-noise in older adults. *Neurobiology of Aging, 81*, 102–115. <https://doi.org/10.1016/j.neurobiolaging.2019.05.015>.