NeuroImage 175 (2018) 150–160

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Statistical learning of multisensory regularities is enhanced in musicians: An MEG study

Evangelos Paraskevopoulos [a,b,*], Nikolas Chalas [b,c], Panagiotis Kartsidis [a], Andreas Wollbrink [b], Panagiotis Bamidis [a]

[a] School of Medicine, Faculty of Health Sciences, Aristotle University of Thessaloniki, 54124, Thessaloniki, Greece
[b] Institute for Biomagnetism and Biosignalanalysis, University of Münster, D-48149, Germany
[c] School of Biology, Faculty of Science, Aristotle University of Thessaloniki, 54124, Greece

ARTICLE INFO

Keywords:
Multisensory learning
Magnetoencephalography
Statistical learning
Music neuroscience
Plasticity

ABSTRACT

The present study used magnetoencephalography (MEG) to identify the neural correlates of audiovisual statistical learning, while disentangling the differential contributions of uni- and multi-modal statistical mismatch responses in humans. The applied paradigm was based on a combination of a statistical learning paradigm and a multisensory oddball one, combining an audiovisual, an auditory and a visual stimulation stream, along with the corresponding deviances. Plasticity effects due to musical expertise were investigated by comparing the behavioral and MEG responses of musicians to non-musicians. The behavioral results indicated that the learning was successful for both musicians and non-musicians. The unimodal MEG responses are consistent with previous studies, revealing the contribution of Heschl's gyrus for the identification of auditory statistical mismatches and the contribution of medial temporal and visual association areas for the visual modality. The cortical network underlying audiovisual statistical learning was found to be partly common and partly distinct from the corresponding unimodal networks, comprising right temporal and left inferior frontal sources. Musicians showed enhanced activation in superior temporal and superior frontal gyrus. Connectivity and information processing flow amongst the sources comprising the cortical network of audiovisual statistical learning, as estimated by transfer entropy, was reorganized in musicians, indicating enhanced top-down processing. This neuroplastic effect showed a cross-modal stability between the auditory and audiovisual modalities.

Introduction

The human perceptual system, constantly receiving multisensory input, decides which sensory information to integrate on the basis of stimulus characteristics (Parise and Spence, 2009), prior experiences (Gau and Noppeney, 2016), or learned associations (Paraskevopoulos et al., 2012). In terms of brain activity, this decision relies on the interplay of distributed cortical regions, including the ones traditionally considered as unisensory (Driver and Noesselt, 2008).

Multisensory training constitutes one of the most effective learning mechanisms, inducing effects superior to unisensory training (Ladan Shams and Seitz, 2008). Any association of multisensory stimuli reinforces learning; nonetheless, congruency of the unisensory information further improves the outcome (Kupers et al., 2011). The notion of supramodal brain areas, involved in a particular form of information processing independently from the sensory modality (Rosenblum et al., 2017), may account for the advantage of multisensory learning when the stimuli are congruent. This advantage, is framed physiologically by superadditivity, which indicates that multisensory stimulation causes greater neuronal activation than the sum of the corresponding unisensory stimuli (Stanford and Stein, 2007). Importantly, the effects of multisensory leaning can be generalized at a unisensory level (Holmes and Spence, 2005) and show cross-modal transfer (Fujisaki et al., 2004).

Statistical learning refers to the human ability to segment a stream of sensory information into chunks according to their transitional probabilities (Saffran et al., 1996; Saffran et al., 1999). It provides an implicit mechanism by which the cognitive system extracts the underlying structure of regularities (Rodríguez-Fornells et al., 2009), independently of the input's sensory modality (Mitchel and Weiss, 2011). Frost et al. (2015) proposed that statistical learning is comprised by neural underpinnings that grounds on domain-general functionalities, constrained to operate within specific modalities. Hence, it may be considered as a

* Corresponding author. School of Medicine, Faculty of Health Sciences, Aristotle University of Thessaloniki, 54124, Thessaloniki, Greece.
E-mail address: parasvag@auth.gr (E. Paraskevopoulos).

https://doi.org/10.1016/j.neuroimage.2018.04.002
Received 24 January 2018; Received in revised form 21 March 2018; Accepted 2 April 2018
Available online 3 April 2018

supramodal learning mechanism, respecting restrictions determined by modality-specific networks. The cortical regions involved in auditory statistical learning include temporal sources, inferior frontal gyrus (IFG), and inferior parietal cortices, (Karuza et al., 2013), while visual statistical learning correlates to the lateral occipital cortex, right intraparietal sulcus and inferior temporal gyrus (Turk-Browne et al., 2009; Turk-Browne et al., 2010). Crossmodal associations are proposed to be formed by binding the activity of unimodal sensory areas via polymodal regions such as the superior temporal gyrus (Tanabe et al., 2005).

Long-term musical training, substantially modifies brain function and structure enhancing implicit learning of auditory regularities (Herholz et al., 2009). Thereby, musical expertise enhances auditory statistical learning (François and Schön, 2011). Recent magnetoencephalographic studies by our group (Paraskevopoulos et al., 2011) revealed that musicians, compared to non-musicians, showed increased auditory cortex activity after the onset of sounds violating their transitional probabilities, (Paraskevopoulos et al., 2011), as well as enhanced functional connectivity between cortical regions related to statistical learning (Paraskevopoulos et al., 2017). Relevant cross-sectional (Mandikal Vasuki et al., 2017; Schön and François, 2011) and longitudinal (Schön and François, 2011) electroencephalographic studies also documented the effect of musical training. Likewise, musical expertise enhances audiovisual integration (Lee and Noppeney, 2011; Nichols and Grahn, 2016), even for associations based on newly learned rules (Paraskevopoulos et al., 2012). Nonetheless, it is still unclear whether musical training, as a multisensory form of training, enhances statistical learning of audiovisual patterns and how this is related to musicians' enhanced performance in unisensory statistical learning.

The present study investigates the influence of musical expertise on statistical learning of audiovisual stimuli. To this aim, we used MEG measurements to study cortical responses of musicians and non-musicians on a multisensory statistical learning paradigm, which incorporated three parallel streams of stimulation with independent transitional probabilities: an auditory, a visual, and an audiovisual one. The stimulation stream included patterns independently violating the transitional probabilities of each modality, allowing us to assess the differential contribution of the cortical regions involved on multi-vs. unisensory statistical learning. Our hypothesis is that musicians will show stronger cortical activation and reorganized connectivity in audiovisual statistical incongruency responses and that this response will be differentiated from the uni-sensory statistical mismatch responses (Koelsch et al., 2016).

Methods

Subjects

The sample of the present study consisted of 24 subjects, 12 musicians and 12 controls. Musicians (mean age = 26.33; *SD* = 4.0; 4 males) were students of the Music Conservatory in Münster (mean musical training = 15.47; *SD* = 3.72). The controls were non-musicians (mean age = 26.9; *SD* = 5.74; 4 males) and had not received musical education apart from any compulsory school lessons. All subjects were evaluated by clinical audiometry and had normal hearing. Additionally, all subjects were right handed according to the Edinburgh Handedness Inventory (Oldfield, 1971). Subjects provided informed consent in written form prior to their participation in the study. The study protocol was approved by the ethics committee of the Medical Faculty of the University of Münster and the study was conducted according to the Declaration of Helsinki.

Stimuli

Each presented stimulus had 4 distinct characteristics: shape, color, pitch and timbre. These characteristics were combined in order to build up 3 concurrent but independent and parallel information streams: an audiovisual (based on the combination of shape and pitch), an auditory (based on timbre) and a visual (based on color). Sounds were generated via GarageBand (version 2.2; Apple inc.) with a sampling rate of 48 kHz, 16bit. An Attack-Decay-Sustain-Release (ADSR) Envelope was applied in each tone, producing a sound with duration of 400 ms, including 30 ms rise and decay time; equally normalized with Peak Loudness normalization method, as applied by WavePad Sound Editor (version 7.12; NCH software). The auditory stream included 11 different timbres each one related to a specific pitch. The different pitches used were: C, C#, D, D#, E, F, F#, G, G# A, B. The different timbres included: reflective string; pop flute; tenor sax; grand piano; fingerstyle bass; future flute; swirling piano; smokey clav; pop organ; hollywood strings; electric tremolo. The visual stream included 11 different colors, each one belonging to an arbitrary complex shape. The different colors used the following RGB values: (192, 0, 0; 160, 81,16; 132, 140, 142; 255, 192, 0; 175, 170, 105; 0, 176, 80; 70, 181, 211; 173, 173, 219; 0, 32, 96; 112, 48, 160; 127, 127, 127). Each of the shapes had the same maximum height and width and was presented on the center of a black background. The Inter Stimulus Interval (ISI) was 150 ms and it was embedded within each pattern as well as across the patterns.

The stimuli where combined in patterns of 3, respecting the statistical learning paradigm as proposed by Saffran et al. (1999). The audiovisual stream included 11 tones, which were coupled with the 11 different shapes building up the audiovisual associations. Pitch patterns and shapes were adapted from Saffran et al. (1999) and Fiser and Aslin (2001), respectively.

Four different categories of patterns were prepared: standard, audiovisual incongruent, auditory deviant and visual deviant. Each category included 6 patterns. Each standard pattern included 3 stimuli incorporating the correct associations between shape, color, pitch and timbre while at the same time respected the statistical regularities of stimulus patterns of all independent streams (i.e. audiovisual, auditory and visual). Each audiovisual incongruent pattern included 3 stimuli each incorporating the correct associations between color and timbre, being standard with respect to the transitional probabilities of the auditory and visual stream, while the last of the three stimuli had incorrect association of shape and pitch, being incongruent with respect to the statistical regularity of the audiovisual stream only. The shape and pitch association used in the last stimulus of the audiovisual incongruent pattern was taken from the first stimulus of another standard pattern.

Similarly, each auditory deviant included 3 stimuli each incorporating the correct associations between color, pitch and shape, being standard with respect to the transitional probabilities of the audiovisual and visual stream, while the last of the three stimuli had incorrect timbre, producing an auditory deviant. The timbre used in the last stimulus was taken from the first stimulus of another standard pattern. Each visual deviant pattern included 3 stimuli each incorporating the correct associations between timbre, shape and pitch, while the last of the three stimuli had incorrect color, violating the statistical regularities of the visual stream. The color used in the last stimulus was taken from the first stimulus of another standard pattern. An abstract description of the stimuli patterns is depicted in Fig. 1, while a detailed description of all stimuli used is presented in the Supplementary Fig. 1.

MEG recordings - instrumentation

Evoked magnetic fields were recorded in a magnetically shielded room via a 275 channel whole-head system (OMEGA, CTF Systems Inc, Port Coquitlam, Canada). Data were continuously acquired using a sampling rate of 1200 Hz. Subjects were seated upright, and their head position was comfortably stabilized inside the MEG dewar using pads. The visual stimuli were projected onto the back of a semi-transparent screen positioned approximately 90 cm in front of the subjects' nasion using an Optoma EP783S DLP projector with a refresh rate of 60 Hz. The viewing angle ranged from -1.15° to 1.15° in the vertical direction and from -3.86° to 3.86° in the horizontal direction. Auditory stimuli were

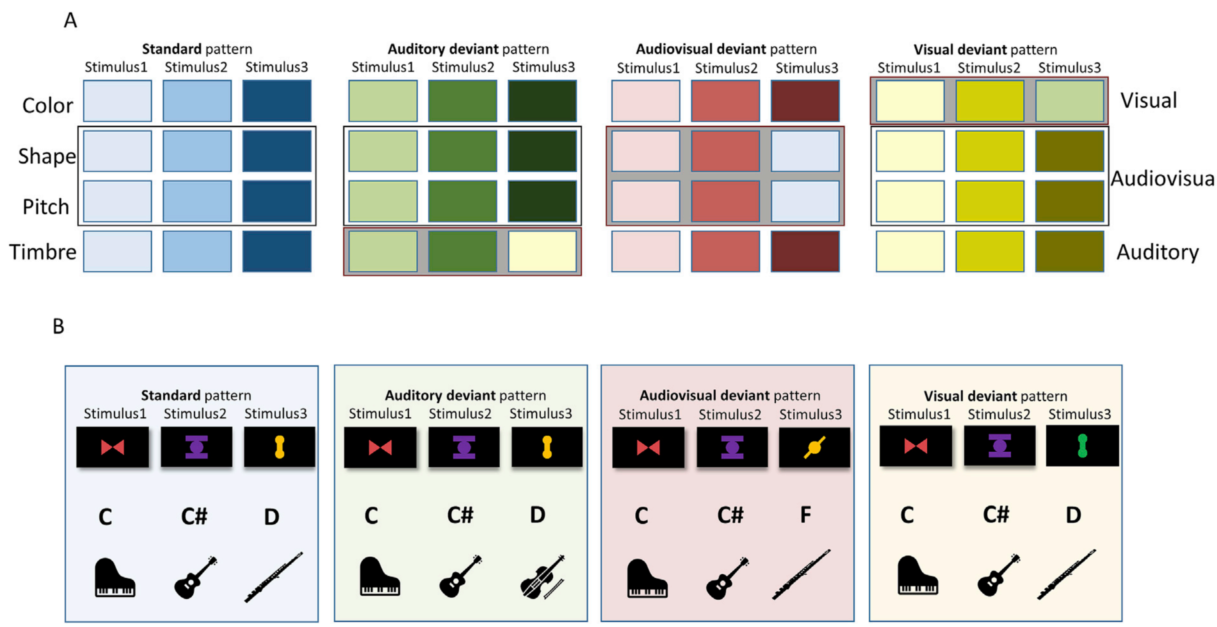

Fig. 1. *Illustration of the design.* A: Each column of squares represents a stimulus. Each row a characteristic: the color, the shape, the pitch and the timbre. Each color represents one pattern, while the brightness of the color represents the position of each stimulus in the pattern. The auditory deviant pattern differs from the standard one in the last stimulus, by using a timbre, which corresponds to the first stimulus of another standard pattern, while keeping all other characteristics correct. The audiovisual incongruent pattern differs from the standard one in the last stimulus, by using a shape and pitch, which corresponds to the first stimulus of another standard pattern, while keeping timbre and color correct. The visual deviant pattern differs from the standard one in the last stimulus, by using a color, which corresponds to the first stimulus of another standard pattern, while keeping all other characteristics correct. B: Example of stimuli representing the design as depicted in Fig. 1A.

delivered via 60 cm long silicon tubes at 60 dB SL above the individual hearing threshold, which was determined with an accuracy of 5 dB at the beginning of each MEG session for each ear. The subject's alertness and compliance were verified by video monitoring. The subjects listened to the three blocks with short breaks in between.

Experimental design

The experimental design consists of a typical statistical learning task, as proposed by Saffran et al. (1999) appropriately adapted for a multisensory oddball paradigm (Paraskevopoulos et al., 2012). Specifically, the paradigm consists of three phases. In the initial phase, which followed the principals of a typical statistical learning task, 70 patterns from the standard category were randomly interleaved and presented. This run established the representations of the upcoming statistical regularities and had a duration of 1.94 min. The transitional probabilities in this phase ranged between .31 and 1. In the following phase (with no break between the presentation of these two phases), the 4 different stimulus pattern categories (i.e. standard, audiovisual incongruent, auditory deviant, visual deviant) were equally and randomly interleaved in a multisensory oddball paradigm. It has to be noted here, that following the principles of the multi-feature Mismatch Negativity Paradigm (Näätänen et al., 2004), the deviants of one modality serve as standards for the rest of the modalities, and hence, the total standard to deviant ratio was 0.25, independently for each modality. This paradigm had 2 constraints: a) at least 1 standard pattern had to occur between the presentation of two deviant patterns and b) the same pattern could not occur in two successive trials. The transitional probabilities in this phase ranged from 0.23 to 0.75 for the standards and from 0.07 to 0.25 for the deviants. The subjects underwent 3 runs of this multisensory oddball paradigm, each one lasting 9.5 min and consisting of 460 patterns. Information given to the subjects prior to the presentation of these two phases included only that they will see some shapes and listen to some sounds and that they should try to keep their attention to the presented stimuli.

Subsequently, participants were informed about a surprise behavioral test phase. The behavioral assessment was performed in the same room as the MEG measurement immediately after the MEG recording. Subjects continued sitting in the same position while listening to the stimuli and answered via a button presses. A test containing 36 pairs of one standard and one deviant pattern was conducted. The deviant's sensory modality (i.e. auditory, visual and audiovisual) was counterbalanced across the 36 pairs, having 12 trials for each modality. The stimulus patterns in each trial were separated by 300 ms and the inter-trial interval was 3 s. The order of the standard and deviant pattern within each pair was also counterbalanced. Subjects had to indicate which of the two patterns of each pair was more familiar to them.

MRI protocol

Weighted T1-MR images from each individual were obtained prior to the experiment, in a 3-T scanner (Gyroscan Intera T30, Philips, Amsterdam, Netherlands). Individualized Finite Element Models (FEM) were constructed and used in the source reconstruction. 400 contiguous T1-weighted slices of 0.5 mm thickness in the sagittal plane (TR = 7.33.64 ms, TE = 3.31 ms) were collected by a Turbo Field Echo acquisition protocol. The field of view was set to 300 × 300 mm with an in-plane matrix of 512 × 512 setting defining the native voxel size at 0.5 × 0.58 × 0.58 mm^3. The intensity bias of the images was then regularized using SPM8 (Statistical Parametric Mapping, http://www.fil.ion.ucl.ac.uk), in order to account for intensity differences within each tissue and the images were resliced to isotropic voxels of 1.17 × 1.17 × 1.17 mm.

Behavioral data analysis

The percentage of correct and incorrect responses for each subject and each modality were calculated, in order to behaviorally evaluate statistical learning on the basis of the applied 2-alternative-force-choice (2AFC) task. As the responses to a 2AFC task originate from a binomial distribution, in order to perform parametric statistics, the percentages were transformed via a logit transformation (Baum, 2008; Fleiss et al., 2013). A one sample *t*-test for each group, comparing the logit transformed percentage of correct responses against the chance level was initially performed. Additionally, an independent samples *t*-test between the groups compared the logit transformed percentage of correct responses between musicians and non-musicians. Last, a mixed model ANOVA with within subjects factor modality (audiovisual, auditory and visual) and between subjects factor group (musicians and non-musicians), was conducted in order to evaluate whether the behavioral effects of the different modalities were differentiated across groups. All analyses were performed using SPSS 23 software (SPSS Inc., Chicago, IL, USA) and the significance level was set to $p < 0.05$.

MEG data analysis

Source activity estimation

Pre-processing of the MEG data was performed using the software Brain Electrical Source Analysis software (BESA research, version 6, Megis Software, Heidelberg, Germany). An adaptive artifact-correction (Ille et al., 2002) method was applied to correct for artifacts due to blinks or eye movements. The recorded data were separated in epochs of 700 ms including a pre-stimulus interval of 150 ms. Epochs were baseline corrected using the interval from −100 to 0 ms. Data were filtered offline with a high pass forward filter of 1 Hz, a low pass zero-phase of 30 Hz and an additional notch filter at 50 Hz. Epochs containing signals larger than 2.5 pT in the MEG were considered artifact contaminated and excluded from averaging. Measurements of all three blocks were averaged in order to achieve the best signal-to-noise ratio possible. Standards and deviants were separately averaged, while the epoch was synchronized to the last stimulus of each pattern. The subset of standards directly preceding the deviants were used in the averaging of the standards so that the two conditions (standards and deviants) of each modality share a signal to noise ratio and epoch number.

A FEM based on the segmentation of 4 different head tissues [i.e. scalp, skull, cerebrospinal fluid (CSF) and brain] was computed for each participant on the basis of their individual structural MRI, and used as a volume conductor model for the source reconstruction. The FEM models were constructed using BESA MRI utility. The individual positions MEG sensors were coregistered via anatomical landmarks (nasion, left and right preauricular points) to each subject's structural MRI. A 3D-spline interpolation of the original MRI was use to transform to ACPC- and Talairach space. The conductivity value for the skin compartment was set to 0.33 S/m, for the skull compartment to 0.0042 S/m, for the brain to 0.33 S/m and for the CSF to 0.79 S/m as proposed by (Geddes and Baker, 1967). This procedure has been shown to improve the source localization reliability of EEG and MEG signals (Wolters et al., 2006), especially when the compartment of CSF is included in the model (Cho et al., 2015).

Current density reconstructions (CDR) were calculated on the neural responses of each subject for each stimulus category (audiovisual congruent, audiovisual incongruent, auditory deviant, visual deviant) using the LORETA method (Pascual-Marqui and Michel, 1994). LORETA has been previously used successfully for the mapping audiovisual incongruences (Paraskevopoulos et al., 2014a; Paraskevopoulos et al., 2012) and has the advantage of not needing an a priori definition of the number of activated sources. A time window of 40 ms was used for the CDR (85–125 ms). The appropriate time window was determined so as to include the rising slope and peak of the N1 component for both groups and all conditions in the grand average Global Field Power. The images were smoothed and their intensities normalized by convolving an isotropic Gaussian kernel with 7 mm full width half-maximum (FWHM) through BESA's smoothing utility.

Statistical analysis of MEG data

The cortical localization of the incongruency response for each modality was defined via the statistical analysis of the MEG data. Statistical analysis of the CDRs was performed using SPM8 and GLM-Flex (http://mrtools.mgh.harvard.edu/index.php?title=Main_Page) running on Matlab (Math Works Inc., Natick, MA, USA). Specifically, using GLM-Flex a separate analysis was designed for each modality (audiovisual, auditory and visual) to create a 2 × 2 mixed model ANOVA with between subject factor group (Musicians and non-musicians) and within subject factor condition (standard and deviant). We tested the main effect of condition (mismatch response) and the group × condition interaction for each modality. Additionally, two 2 × 2 × 2 mixed model ANOVAs with within subjects factors modality (audiovisual and auditory/audiovisual and visual), condition (standard and deviant) and between subjects factor group (musicians and non-musicians) were conducted in order to compare the effect of long-term music training across the different modalities. Results were constrained in gray matter using a mask, thereby keeping the search volume small and in physiologically reasonable areas. Family Wise Error (FWE) correction at 5% level was applied for this whole head analysis, effectively controlling for multiple comparisons, unless otherwise noted.

Connectivity analysis

The definition of the cortical network was performed via the estimation of TE between the source waveforms, in order to analyze the corresponding pattern of the information flow between the cortical regions involved in audiovisual statistical learning. An equivalent current dipole model based on the results of the source activity estimation was used to discriminate between regional brain activity and system noise or uncorrelated activity (Tesche et al., 1995). One dipole was fitted at the peak of each cluster of statistically significant activation, as found in the interaction of the statistical analysis of the source activity for the audiovisual modality, resulting in a 4 dipole model. The coordinates of the dipoles were determined by each cluster's peak coordinates while the orientation was fitted individually in the FEM volume conductor using the averaged data across the standards and audiovisual incongruent trials. The waveforms of activity for the complete epoch of each cortical region involved in audiovisual statistical learning were obtained via source space projection (Tesche et al., 1995).

The Matlab toolbox HERMES (Niso et al., 2013) was used for calculating the 4 × 4 adjacency matrix from dipoles' waveforms for each subject and each condition based on the algorithm of Transfer Entropy (TE). TE is a non-parametric statistic measuring the amount of directed (time-asymmetric) transfer of information between two random processes. The main advantage of TE is that, being based on probability distributions, it detects higher order correlations. Therefore, its result is not dependent on any specific model of the data (Niso et al., 2013). The Network Based Statistic (NBS) (Zalesky et al., 2010) toolbox was used to identify statistically significant connections in the networks. Specifically, a 2 × 2 mixed model ANOVA with between subject factor group (musicians and non-musicians) and within subject factor condition (standard and deviant), was designed to explore the group × condition interaction. The significance level was set to $p < 0.001$ corrected for multiple comparisons via FDR correction.

Results

Behavioral responses

In order to test whether audiovisual statistical learning is reflected behaviorally, the performance of in the 2AFC test was compared to the chance level. This analysis revealed that, both group's performance differed significantly from the chance level [musicians: $t\,(11) = 6.528$, $p < .001$; non-musicians: $t\,(11) = 4.220$, $p < .001$; one – sample *t*–test] indicating that the participants learned to explicitly distinguish between the standard and deviant (congruent and incongruent) patterns on the behavioral level, independently of the modality. Moreover, the two groups (musicians and non-musicians) differed significantly in the accuracy of their responses [$t\,(22) = 2.477$, $p < .05$; independent sample *t* – test], indicating that musicians (mean correct responses after logit transformation: 0.9223; $SD = 0.489$) scored significantly higher than non-musicians (mean correct responses after logit transformation: 0.4774; $SD = 0.407$). The group averages of the percentage of correct responses are shown in Fig. 2. The interaction of modality × group for the standard stimuli shows no significant result [$F(2, 22) = 0.175$; $p = .840$], indicating that at the behavioral level, the effect of statistical learning is equally distributed across the different modalities. Additionally, this result indicates that, at the level of behavior, musical expertise affects equally all three modalities. For the group of musicians, the mean percentage of correct responses in the audiovisual modality = 0.87; $SD = 0.128$; the mean percentage of correct responses in the auditory modality = 0.89; $SD = 0.055$; and the mean percentage of correct responses in the visual modality = 0.85; $SD = 0.122$. While for the group of non-musicians, the mean percentage of correct responses in the audiovisual modality = 0.88; $SD = 0.053$; the mean percentage of correct responses in the auditory modality = 0.87; $SD = 0.056$ and the mean percentage of correct responses in the visual modality = 0.85; $SD = 0.065$.

MEG source activity

The grand average of the Global Field Power (GFP) of each condition and for each group was calculated, as a gross index of the overall activity (Fig. 3). The main effect of condition of the statistical analysis for the audiovisual modality was calculated in order to reveal the cortical responses to incongruent trials, which violated the statistical regularities of the audiovisual stream of stimulation. This analysis revealed 2 clusters which showed significantly stronger activity in the incongruent than the congruent condition, indicating that these regions underpin the identification of incongruent regularities violating the statistical properties of the audiovisual part of the stream. One cluster was located on the right Heschl's gyrus and one on the left inferior frontal gyrus. Peak coordinates and statistical values for the activations are presented in Table 1, while the statistical map is presented in Fig. 4.

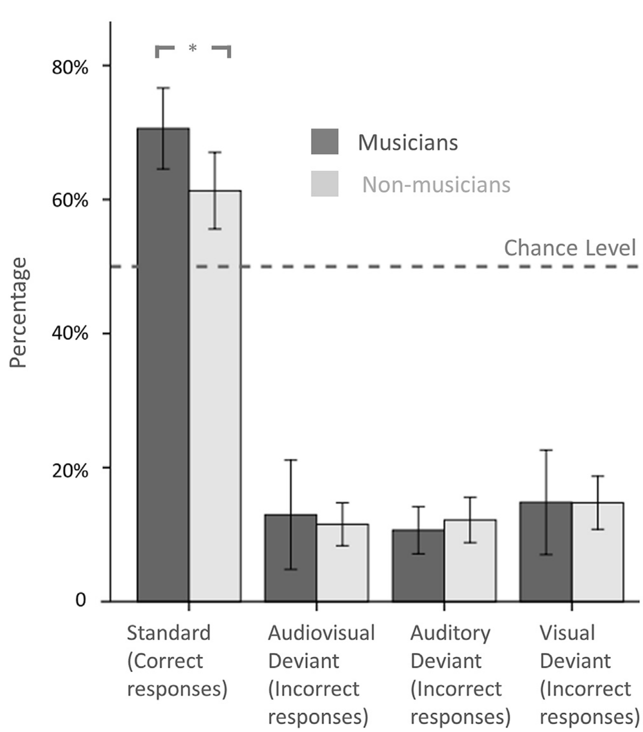

Fig. 2. *Behavioral responses.* Percentage of correct and incorrect behavioral responses, for musicians and non-musicians, in the statistical learning 2-alternative-forced-choice test. The dashed line indicates the chance level. Error bars indicate 95% confidence intervals.

The interaction of group × condition was calculated in order to evaluate whether the group of musicians differs significantly from the group of non-musicians in the identification of incongruent audiovisual regularities. This analysis showed a statistically significant effect, also located in 2 clusters: one on the right superior temporal gyrus and one on the right superior frontal gyrus in the pre-supplementary motor area (pre-SMA). Thereby it was revealed that musicians showed significantly greater activity in these regions when confronted to an audiovisual statistical incongruency. Peak coordinates and statistical values are presented in Table 2, while the statistical map is presented in Fig. 5.

The main effect of condition of the statistical analysis for the auditory modality was calculated in order to reveal the cortical responses to deviant trials, which violated the statistical regularities of the auditory stream of stimulation. This analysis revealed one cluster which showed significantly stronger activity in the deviant than the standard pattern. This cluster was located on the right Heschl's gyrus. Peak coordinates and statistical values for the activations are presented in Table 1, while the statistical map is presented in Fig. 4.

In order to evaluate whether the group of musicians differs significantly from the group of non-musicians in the identification of deviant auditory regularities, we calculated the interaction of group × condition for the auditory modality. This analysis showed a statistically significant effect, located in 4 clusters: one on the right superior temporal gyrus, one on the right superior frontal gyrus in the pre-supplementary motor area, one on the left precentral gyrus and one on the right inferior frontal gyrus. Thereby it was revealed that musicians showed significantly greater activity in these regions when confronted to an auditory statistical mismatch. Peak coordinates and statistical values are presented in Table 2, while the statistical map is presented in Fig. 5.

The main effect of condition of the statistical analysis for the visual modality was calculated in order to reveal the cortical responses to deviant trials, which violated the statistical regularities of the visual stream of stimulation. This analysis revealed several clusters, which showed significantly stronger activity in the standard than the deviant condition, in contrast to the modalities previously mentioned. Two of these clusters were located on the right middle temporal gyrus. Another cluster was located on the left middle temporal gyrus, one on the left middle occipital gyrus and one on the precuneus. Peak coordinates and statistical values for the activations are presented in Table 1, while the statistical map is presented in Fig. 4. The interaction of group × condition yielded no significant results.

Results of the interaction of modality × condition × group show that there is a significant difference amongst the different groups in the identification of deviance, between the audiovisual and the auditory modality located on the right middle superior temporal gyrus. Additionally, a significant difference between the different groups in the identification of deviance amongst the audiovisual and the visual modality was located in 2 clusters: one at on the superior temporal gyrus along and one on the pole of the frontal lobe, on the superior frontal gyrus. The statistical maps of the modality × condition × group interaction are presented in Fig. 6. Peak coordinates and statistical values for the activations are presented in Table 3.

Lastly, in order to relate the behavioral and the neural data, we performed an ANCOVA with between subjects factor Group (musicians and non-musicians), within subjects variable the cortical responses to the standard condition and entered the behavioral performance of the subjects as a covariate. Results indicated an interaction of the covariate and the group, located in the right middle temporal gyrus (peak coordinates: $x = 60$, $y = -6$, $z = 16$; $t\ (22) = 2.475$; cluster size = 285 voxels; $p = 0.011$ uncorrected). This result indicates that the greater the accuracy in the behavioral responses, the greater the activation in this region; and that this effect is even stronger for the group of musicians. Nonetheless, this result is significant only at an uncorrected level and should be taken into account cautiously.

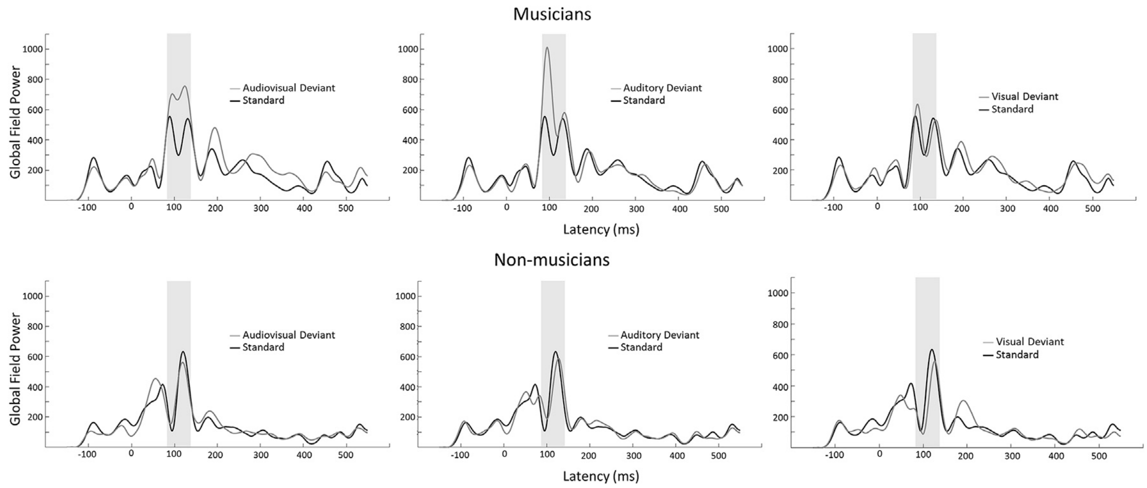

Fig. 3. *Global Field Power.* The grand average global field power for musicians (top row) and non-musicians (bottom row) in the standard (black line) and deviant (gray line) responses for the audiovisual (left side) auditory (middle) and visual (right side) modalities. Global Field Power is depicted in order to reveal the overall sensor activation pattern. The time interval where the analysis was performed is marked gray.

Table 1
Location, peak voxel statistical value and cluster size of MEG source activity for the main effect of condition (deviant > standard) for the audiovisual and the auditory modality and the main effect of condition (deviant < standard) for the visual modality.

Modality	Location of activation	Coordinates			Peak voxel *t* (22) value	*P* value FEW - Corrected	Cluster size
		X	Y	Z			
Audiovisual	Right Heschl's Gyrus	42	−18	9	−6.474	0.001	1053
	Left Inferior Frontal Gyrus	41	34	−14	−5.63	0.001	879
Auditory	Right Heschl's Gyrus	−41	−17	6	−5.064	0.001	2043
Visual	Right Middle Temporal Gyrus	69	−32	−4	5.18	0.001	359
	Right Middle Temporal Gyrus	45	−66	6	5.023	0.001	849
	Left Middle Temporal gyrus	−51	−30	2	5.056	0.001	189
	Left Middle Occipital Gyrus				5.229	0.001	409
	Precuneus	15	−78	46	5.249	0.001	598

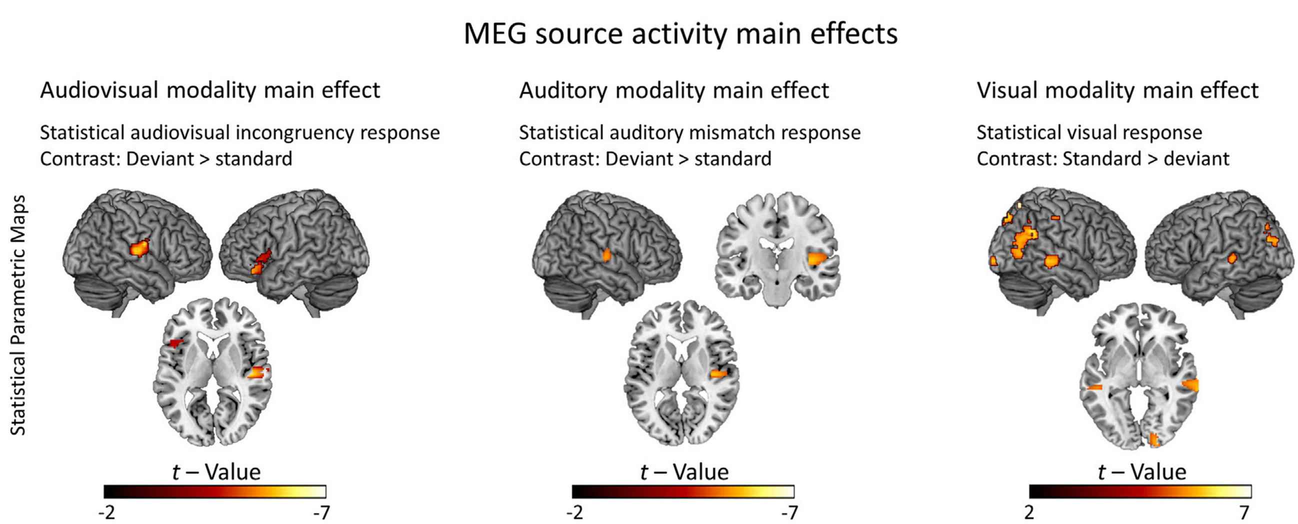

Fig. 4. *MEG source activity main effects.* Statistical parametric maps of the main effects of condition (deviant > standard) for the audiovisual and the auditory modality and the main effect of condition (deviant < standard) for the visual modality as revealed by the Flexible Factorial Model. Threshold: FWE corrected at $p < 0.05$.

Cortical network analysis results

The network analysis, conducted in order to identify differences between the groups in the information flow between the cortical regions involved in audiovisual statistical learning, revealed that all nodes of the network are connected amongst them, creating a complete graph. As this is the interaction effect of the analysis, this result indicates that musicians and non-musicians show differences in the information flow between the

Table 2
Location, peak voxel statistical value and cluster size of MEG source activity for the interaction of condition (deviant ≠ standard) × group (musicians ≠ non-musicians) of each modality.

Modality	Location of activation	Coordinates			Peak voxel F (1, 22)	P value FWE- Corrected	Cluster size
		X	Y	Z			
Audiovisual	Right Superior Temporal Gyrus	43	−24	−2	40.920	0.001	591
	Right Superior Frontal Gyrus	34	−11	68	27.271	0.001	4291
Auditory	Right Superior Temporal Gyrus	47	−23	0	21.76	0.001	515
	Right Superior Frontal Gyrus	29	−9	62	23.313	0.001	678
	Left Precentral Gyrus	−31	−10	46	21.765	0.001	425
	Right Inferior Frontal Gyrus	55	12	18	14.916	0.008	252

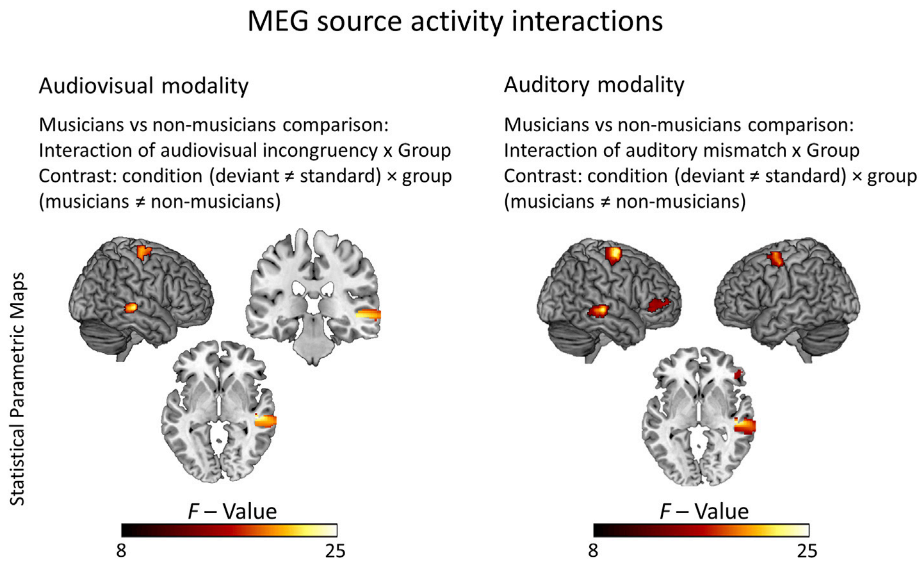

Fig. 5. *MEG source activity interactions.* Statistical parametric maps of the interaction of condition (deviant ≠ standard) × group (musicians ≠ non-musicians) for each modality as revealed by the Flexible Factorial Model. Threshold: FWE corrected at $p < 0.05$.

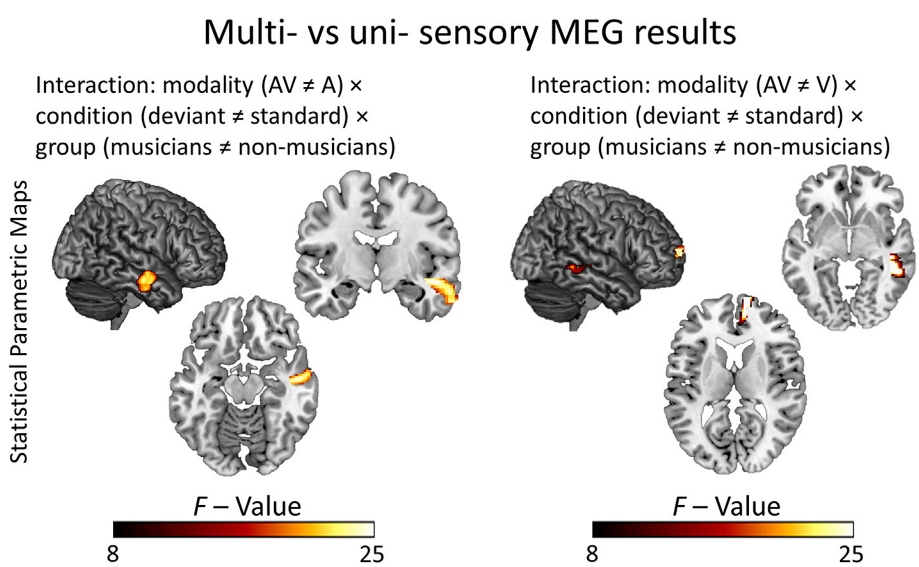

Fig. 6. *Multi-vs. Uni-sensory MEG results.* Statistical parametric maps of the modality (AV ≠ A and AV ≠ V) × condition (deviant ≠ standard) × group (musicians ≠ non-musicians) interactions as revealed by the Flexible Factorial Model. Threshold: FWE corrected at $p < 0.05$.

corresponding nodes, when processing statistical learning patterns. The direction of the flow of information, as calculated via TE, indicates that musicians and non-musicians show statistically significant connectivity differences in the strength of the information that the right Heschl's gyrus receives (i.e. incoming edges) from right superior temporal gyrus, the right superior frontal gyrus and the left inferior frontal gyrus. Additionally, musicians and non-musicians show significant connectivity differences in the information that the right superior temporal gyrus receives

Table 3
Location, peak voxel statistical value and cluster size of MEG source activity for the interaction of modality (AV ≠ A and AV ≠ V) × condition (deviant ≠ standard) × group (musicians ≠ non-musicians).

Modality factor conditions	Location of activation	Coordinates			Peak voxel *F* (1, 22)	*P* value FWE- Corrected	Cluster size
		X	Y	Z			
Audiovisual × Auditory	Right Middle Temporal Gyrus	60	−6	20	16.1157	0.007	416
Audiovisual × Visual	Superior Frontal Gyrus	10	58	10	24.1932	0.001	547
	Superior Temporal Gyrus	48	−40	−4	19.464	0.002	724

from the right superior frontal gyrus as well as the information that this region sends to the right Heschl's gyrus and the left inferior frontal gyrus. Significant connectivity differences between the two groups are also revealed in the information that the left inferior frontal gyrus receives from the right superior temporal gyrus and the right superior frontal gyrus, as well as in the information that that the left inferior frontal gyrus sends to the right Heschl's gyrus. The right superior frontal gyrus (pre-SMA) showed significant connectivity differences between musicians and non-musicians in the strength of the outgoing connections to all other nodes in the network, i.e. right Heschl's gyrus, right superior temporal gyrus, and left inferior frontal gyrus. Within this context, sending information may also be interpreted as modulating the activity of the other area. The significance threshold for this analysis was set to $p < 0.001$ corrected for multiple comparisons via FDR correction.

An analysis of the network's node strength was also conducted in order to identify the importance of each node in the network. Node strength is the sum of the weights (i.e. *F* values in the present analysis) of incoming and outgoing edges connected to the node, hence, identifying the nodes with greater connectivity differences. This analysis showed that the node with greater node strength was superior temporal gyrus, followed by Heschl's gyrus and inferior frontal gyrus while the node with the least strength was the right pre-SMA. The statistical map of the network analysis depicting the interaction of group × condition for the audiovisual modality along with the node strength analysis are depicted in Fig. 7.

In order to identify the origin of this interaction we also analyzed the corresponding networks within each group. These analyses did not yield statistically significant networks, probably due to the smaller sample. Nonetheless the results showed that the mean information flow between all of the nodes of the network was increased for the group of musicians, compared to non-musicians, thereby indicating that the significant interaction effect, as described above, originates from increased connectivity in the group of musicians.

Lastly, we also performed an analysis of the interaction of modality (audiovisual vs. visual) × condition (standard vs. deviant) × group (musicians vs. non-musicians), in order to estimate the functional connectivity of the regions identified as statistically significant in the corresponding MEG source activity analysis. The result of this interaction showed no significant effect. The network analysis of modality (audiovisual vs. auditory) × condition (standard vs. deviant) × group (musicians vs. non-musicians), was not performed as the corresponding MEG source activity results revealed only one cluster of activity, and hence, there was no reason to estimate functional connectivity.

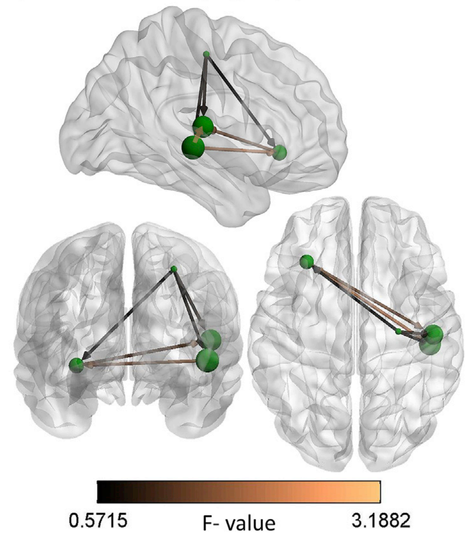

Fig. 7. *Differences in connectivity between musicians and non-musicians.* Statistical parametric maps of the significant networks for the interaction of the factors group (musicians ≠ non-musicians) × condition (deviant ≠ standard) in the audiovisual modality. The network presented is significant at $p < 0.001$, FDR corrected level. The color scale of the edges indicates *F* values, while the size of the node represents node strength.

Discussion

The present study used MEG to identify the neural correlates of audiovisual statistical learning, while disentangling the differential contributions of uni- and multi-modal statistical mismatch responses. The functional connectivity of the cortical network supporting audiovisual statistical learning was investigated by means of a statistical comparison of the estimated transfer entropy (TE) in the sources' activity. Plasticity effects of musical expertise were investigated by comparing musicians to non-musicians. Results indicated that the cortical network underlying audiovisual statistical learning, involving right temporal and left inferior frontal sources, is partly common and partly distinct from the unimodal networks. Musicians, compared to non-musicians, showed increased activity in the frontal and temporal areas contributing to the audiovisual statistical incongruency and the auditory statistical mismatch response, but did not show significant differences in the processing of visual statistical mismatches. Moreover, the information flow pattern between the regions supporting audiovisual statistical learning was reorganized in the group of musicians, compared to non-musicians, showing enhanced functional connectivity between the left IFG, the right STG and pre-SMA.

The behavioral results of the present study indicate that both groups were able to extract the stimulation's underlying statistical regularity, discriminating between the correct and incorrect patterns presented, independently of the modality. Hence, statistical learning was automatically performed for all three modalities. The resulting representation of the statistical regularities was available for explicit comparison with new input, as indicated by the 2AFC test. This is in line with previous studies of audiovisual statistical learning (Mitchel et al., 2014; Shams et al., 2010). The behavioral results also indicate that musicians' prior experience introduced a behavioral benefit, enhancing statistical learning independently of the modality. This is in line with previous studies investigating musical expertise effects on a variety of different aspects of audiovisual processing (Aizenman et al., 2017; Lee and Noppeney, 2011; Paraskevopoulos et al., 2012), as well as statistical learning of auditory material (Schön and François, 2011). The fact that the behavioral responses do not show an interaction of modality × group, indicating that at a behavioral level musical expertise affects equally all three modalities. Interestingly, the MEG data show a significant result in the corresponding interaction. This may be attributed to the fact that the behavioral responses are responses to a 2-alternative forced choice test which is the result of multiple, diverse processing schemes at different stages (Nieder and Miller, 2003), while the neural responses show cortical processing at one specific time-point, and hence greater sensitivity with regard to this processing stage alone. An increased sensitivity of the neural data in comparison to the behavioral ones was also shown in a previous manuscript of our group regarding statistical learning in the auditory domain (Paraskevopoulos et al., 2011). Additionally, it has to be noted that the number of trials of the 2AFC test for each modality alone was relatively low (i.e. 12), a fact that may also have affected the sensitivity of the test for each modality. Further research should be conducted at a behavioral level using a greater number of trials for each modality. Lastly, the fact that the incorrect patterns included in the 2AFC test were presented during the oddball paradigm as deviants may have made the behavioral testing more conservative, as the subjects may have familiarized with the patterns at some -smaller than with the correct patterns-extent.

The MEG results of the audiovisual modality revealed that 2 cortical sources contributed to the statistical incongruency response. Specifically, the right Heschl's gyrus and the left IFG were found to show significant activity in the main effect of condition for the audiovisual modality. The right Heschl's gyrus has been repeatedly found to correlate with statistical learning (Daikoku et al., 2014; Paraskevopoulos et al., 2011) as well as audiovisual processing, within the context of language processing (Beauchamp et al., 2004; van Atteveldt et al., 2004; van Atteveldt et al., 2006), and semantic congruency (Hein et al., 2007). Taking into account that this region is highly related to acoustic processing (Mathys et al., 2010; Warrier et al., 2009) our result suggests that multisensory interactions, occur not only in higher-level cortical areas, but also in lower level and probably primary sensory areas. This is consistent with a number of recent studies showing multisensory interactions in primary sensory areas (Driver and Noesselt, 2008; van Atteveldt et al., 2014) a finding which is corroborated by animal studies (Kayser et al., 2005; Kayser et al., 2008).

The main effect of audiovisual incongruency in statistical learning also revealed a significant contribution of the left IFG. The activity of the left IFG is correlated with many different processes, including statistical learning of auditory or linguistic sequences (Abla et al., 2008) as well as the processing of audiovisual sequences (Paraskevopoulos et al., 2012; Stekelenburg and Vroomen, 2007). This is also consistent with the role of Broca's area as a region processing supramodally hierarchically structured sequences (Molnar-Szakacs et al., 2005; Tettamanti and Weniger, 2006; Uddén and Bahlmann, 2012). The cross-modal interactions, within the context of statistical learning, reinforce the sharing of perceptual cues across modalities, and these cues operate as priors for the uni-sensory areas, under the predictive coding hypothesis (Stekelenburg and Vroomen, 2015). The left IFG is highly linked with predictive processes (d'Acremont et al., 2013; Iglesias et al., 2013). This interpretation is in line with the view of Frost et al. (2015) regarding the grounding of statistical learning in domain-general learning principles which are constrained to operate within specific modalities. The interaction of the right Heschl's gyrus and the left IFG also indicates dissociable but complementary roles for these regions in the processing of audiovisual statistical incongruences. This corresponds to the recent work of Uno. et al. (2015) which suggests that the degree of congruency drives the differential contribution of IFG and superior temporal sulcus in audiovisual mismatch detection.

The interaction effect of the audiovisual modality reveals plasticity effects that can be attributed to long-term musical training. Increased activation in the group of musicians was found in the right secondary auditory cortex (STG) and the pre-SMA. The enhancement of activity of the right STG in musicians has been well documented (Paraskevopoulos et al., 2014b; Schlaug et al., 2005) also within the context of audiovisual associations (Pantev et al., 2015). The involvement of pre-SMA may be interpreted as a contribution of motor association areas originating from the strong audio-motoric binding that musicians have highly trained (Zatorre et al., 2007). This region has been shown in a previous study by our group to contribute to the identification of abstract audiovisual incogruences (Paraskevopoulos et al., 2012).

The cortical network analysis indicates stronger connectivity in the group of musicians, compared to the group of non-musicians, between the right STG, right Heschl's gyrus, left IFG and right superior frontal gyrus. The enhanced modulation of Heschl's gyrus activity by the superior temporal gyrus is consistent with an enhanced top-down information processing within the auditory cortex in the group of musicians. This interpretation relates to the local hyperconnectivity found in absolute pitch musicians (Loui et al., 2011), even though none of the musicians that were used in the present study had absolute pitch as indexed by self-report. The connectivity between the left IFG and the right temporal sources, with the STG sending information to the left IFG and the left IFG to modulate the activity of the Heschl's gyrus, may be grounded on enhanced inter-cortical connections via the corpus callosum (Schlaug et al., 1995; Steele et al., 2013). This information processing pattern within the context of statistical learning indicates that musicians experience an enhanced top-down processing of audiovisual regularities based on their supramoddally hierarchical structure. Lastly, the right pre-SMA shows enhanced connectivity to the Heschl's gyrus and the STG, probably supported by the enlarged superior longitudinal (Oechslin et al., 2010) and arcuate fasciculi (Halwani et al., 2011) that musicians show, due to their strong audio-motor binding. In a cohesive interpretation of the results for the audiovisual modality, we propose that the primary auditory cortex and the left IFG are responsible for the identification of audiovisual statistical regularities, while the effect of musical training in this process may be described as enhanced connectivity between these regions and additional top-down modulation of their activity by the pre-SMA and the STG.

The main effect of auditory statistical mismatch (Koelsch et al., 2016) was located in the right Heschl's gyrus. This is highly consistent with studies investigating statistical learning of auditory sequences (Daikoku et al., 2014; Paraskevopoulos et al., 2011). The right lateralization indicated by the present results may be a result of the fact that the auditory information carrier was timbre, which repeatedly shows right lateralization in mismatch responses (Paraskevopoulos et al., 2012; Toiviainen et al., 1998). The comparison of auditory statistical mismatches between musicians and non-musicians reveals that musicians show enhanced activity in the right IFG, the right homologue of Broca's area, a result that corresponds to enhanced processing of musical stimuli (Koelsch et al., 2002; Maess et al., 2001). Additionally, musicians show enhanced contribution of pre-SMA, similarly to the audiovisual condition, indicating a cross-modal stability of this neuroplastic effect.

The main effect of visual statistical mismatch was located in bilateral sources in middle temporal and visual association areas. This result is consistent with previous studies in the field of visual statistical learning, highlighting the necessity of medial temporal structures (Schapiro et al., 2014) and the role of category-specific visual regions (Turk-Browne et al., 2009). Interestingly, the two groups did not differ in their visual responses, indicating that the neuroplastic effect of long-term musical training is not generalizable in the visual modality. This may be attributed to a training-task specificity as musical training does not affect color processing, and is consistent with a previous study of our group investigating visual mismatch responses (Paraskevopoulos et al., 2012). Interestingly, the MEG source activity results show that the effect of audiovisual incongruency detection is greater than the effect of the auditory mismatch detection, and that this difference is greater in the group of musicians in for the group of non-musicians. Similarly, the results show that audiovisual incongruency detection is greater than visual mismatch detection, and that this is stronger in the group of musicians. This indicates that the cortical processing of audiovisual regularities is enhanced in musicians and differs significantly from the cortical processing of each of the unisensory modalities, when compared to non-musicians. A probable interpretation of this result may be grounded on superadditive processes (Stanford and Stein, 2007) related to multisensory learning.

Conclusion

The present study used a combination of a typical statistical learning and a multisensory mismatch paradigm to investigate audiovisual statistical learning, while disentangling the differential contributions of uni- and multi-modal responses. Plasticity effects, were investigated by comparing the behavioral and MEG responses of musicians to non-musicians. The unimodal responses indicate contribution of Heschl's gyrus for the identification of auditory statistical mismatches and contribution of medial temporal and visual association areas for visual responses. The cortical network underlying audiovisual statistical learning was partly common and partly distinct from the corresponding unimodal networks, involving right temporal and left inferior frontal sources. Musicians showed enhanced connectivity between these regions and top-down modulation of their activity by the pre-SMA and the STG, while this neuroplastic effect also showed a cross-modal stability between the auditory and audiovisual domain.

Acknowledgements

This work has been supported by a research fellowship from the Greek State Scholarships Foundation (IKY), from the act "Supporting postdoctoral researchers" from grants of the programme "Development of human resources, education and life-long training" with Priority Axes 6,8,9 and co-funded by the European Social Fund - ESF and the Greek State and internal grants of the Institute for Biomagnetism and Biosignalanalysis, University of Münster, Germany. The authors would like to thank Prof. Dr. Christo Pantev for helpful comments on previous versions of this manuscript. We would also like to thank the technicians of the Institute for Biomagnetism and Biosignalanalysis, for supporting data acquisition.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.neuroimage.2018.04.002.

Conflicts of interest

The authors declare no competing financial interests.

References

Abla, D., Katahira, K., Okanoya, K., 2008. On-line assessment of statistical learning by event-related potentials. J. Cognitive Neurosci. 20 (6), 952–964. https://doi.org/10.1162/jocn.2008.20058.

Aizenman, A.M., Gold, J.M., Sekuler, R., 2017, April 29. Multisensory integration in short-term memory: musicians do rock. Neurosci. Pergamon. https://doi.org/10.1016/j.neuroscience.2017.04.031.

Baum, C.F., 2008. Modeling proportions. Stata J. 8 (2), 299–303. https://doi.org/The Stata Journal.

Beauchamp, M.S., Lee, K.E., Argall, B.D., Martin, A., 2004. Integration of auditory and visual information about objects in superior temporal sulcus. Neuron 41 (5), 809–823. https://doi.org/10.1016/S0896-6273(04)00070-4.

Cho, J.-H., Vorwerk, J., Wolters, C.H., Knösche, T.R., 2015. Influence of the head model on EEG and MEG source connectivity analyses. NeuroImage 110, 60–77. https://doi.org/10.1016/j.neuroimage.2015.01.043.

d'Acremont, M., Fornari, E., Bossaerts, P., 2013. Activity in inferior parietal and medial prefrontal cortex signals the accumulation of evidence in a probability learning task. PLoS Comput. Biol. 9 (1) e1002895. https://doi.org/10.1371/journal.pcbi.1002895.

Daikoku, T., Yatomi, Y., Yumoto, M., 2014. Implicit and explicit statistical learning of tone sequences across spectral shifts. Neuropsychologia 63, 194–204. https://doi.org/10.1016/j.neuropsychologia.2014.08.028.

Driver, J., Noesselt, T., 2008. Multisensory interplay reveals crossmodal influences on "sensory-specific" brain regions, neural responses, and judgments. Neuron 57 (1), 11–23. https://doi.org/10.1016/j.neuron.2007.12.013.

Fiser, J., Aslin, R.N., 2001. Unsupervised statistical learning of higher-order spatial structures from visual scenes. Psychol. Sci. 12 (6), 499–504. https://doi.org/10.1111/1467-9280.00392.

Fleiss, J.L., Levin, B., Paik, M.C., 2013. Statistical Methods for Rates and Proportions. Wiley.

François, C., Schön, D., 2011. Musical expertise boosts implicit learning of both musical and linguistic structures. Cereb. Cortex 21 (10), 2357–2365.

Frost, R., Armstrong, B.C., Siegelman, N., Christiansen, M.H., 2015. Domain generality versus modality specificity: the paradox of statistical learning. Trends Cognitive Sci. 19 (3), 117–125. https://doi.org/10.1016/j.tics.2014.12.010.

Fujisaki, W., Shimojo, S., Kashino, M., Nishida, S., 2004. Recalibration of audiovisual simultaneity. Nat. Neurosci. 7 (7), 773–778. https://doi.org/10.1038/nn1268.

Gau, R., Noppeney, U., 2016. How prior expectations shape multisensory perception. NeuroImage 124, 876–886. https://doi.org/10.1016/j.neuroimage.2015.09.045.

Geddes, L.A., Baker, L.E., 1967. The specific resistance of biological material–a compendium of data for the biomedical engineer and physiologist. Med. Biol. Eng. 5 (3), 271–293.

Halwani, G.F., Loui, P., Rüber, T., Schlaug, G., 2011. Effects of practice and experience on the arcuate fasciculus: comparing singers, instrumentalists, and non-musicians. Front. Psychol. 2 (156). https://doi.org/10.3389/fpsyg.2011.00156.

Hein, G., Doehrmann, O., Müller, N.G., Kaiser, J., Muckli, L., Naumer, M.J., 2007. Object familiarity and semantic congruency modulate responses in cortical audiovisual integration areas. J. Neurosci. Official J. Soc. Neurosci. 27 (30), 7881–7887. https://doi.org/10.1523/JNEUROSCI.1740-07.2007.

Herholz, S.C., Lappe, C., Pantev, C., 2009. Looking for a pattern: an MEG study on the abstract mismatch negativity in musicians and nonmusicians. BMC Neurosci. 10 (42). https://doi.org/10.1186/1471-2202-10-42.

Holmes, N.P., Spence, C., 2005. Multisensory integration: space, time and superadditivity. Curr. Biol. CB 15 (18), R762–R764. https://doi.org/10.1016/j.cub.2005.08.058.

Iglesias, S., Mathys, C., Brodersen, K.H., Kasper, L., Piccirelli, M., DenOuden, H.E.M., Stephan, K.E., 2013. Hierarchical prediction errors in midbrain and basal forebrain during sensory learning. Neuron 80 (2), 519–530. https://doi.org/10.1016/j.neuron.2013.09.009.

Ille, N., Berg, P., Scherg, M., 2002. Artifact correction of the ongoing EEG using spatial filters based on artifact and brain signal topographies. J. Clin. Neurophysiology Official Publ. Am. Electroencephalogr. Soc. 19 (2), 113–124.

Karuza, E., Newport, E., Aslin, R., Starling, S., 2013. The neural correlates of statistical learning in a word segmentation task: an fMRI study. Brain Lang.

Kayser, C., Petkov, C.I., Augath, M., Logothetis, N.K., 2005. Integration of touch and sound in auditory cortex. Neuron 48 (2), 373–384. https://doi.org/10.1016/j.neuron.2005.09.018.

Kayser, C., Petkov, C.I., Logothetis, N.K., 2008. Visual modulation of neurons in auditory cortex. Cereb. Cortex 18 (7), 1560–1574 (New York, N.Y. : 1991). https://doi.org/10.1093/cercor/bhm187.

Koelsch, S., Busch, T., Jentschke, S., Rohrmeier, M., 2016. Under the hood of statistical learning: a statistical MMN reflects the magnitude of transitional probabilities in auditory sequences. Sci. Rep. 6, 19741. https://doi.org/10.1038/srep19741.

Koelsch, S., Gunter, T.C.V., Cramon, D.Y., Zysset, S., Lohmann, G., Friederici, A.D., 2002. Bach speaks: a cortical "Language-Network" serves the processing of music. NeuroImage 17 (2), 956–966. https://doi.org/10.1006/nimg.2002.1154.

Kupers, R., Pietrini, P., Ricciardi, E., Ptito, M., 2011. The nature of consciousness in the visually deprived brain. Front. Psychol. 2, 19. https://doi.org/10.3389/fpsyg.2011.00019.

Lee, H.L., Noppeney, U., 2011. Long-term music training tunes how the brain temporally binds signals from multiple senses. Proc. Natl. Acad. Sci. 108 (51), E1441–E1450. https://doi.org/10.1073/pnas.1115267108.

Loui, P., Li, H.C., Hohmann, A., Schlaug, G., 2011. Enhanced cortical connectivity in absolute pitch musicians: a model for local hyperconnectivity. J. Cognitive Neurosci. 23 (4), 1015–1026. https://doi.org/10.1162/jocn.2010.21500.

Maess, B., Koelsch, S., Gunter, T.C., Friederici, A.D., 2001. Musical syntax is processed in Broca's area: an MEG study. Nat. Neurosci. 4 (5), 540–545. https://doi.org/10.1038/87502.

Mandikal Vasuki, P.R., Sharma, M., Ibrahim, R.K., Arciuli, J., 2017. Musicians' online performance during auditory and visual statistical learning tasks. Front. Hum. Neurosci. 11, 114. https://doi.org/10.3389/fnhum.2017.00114.

Mathys, C., Loui, P., Zheng, X., Schlaug, G., 2010. Non-invasive brain stimulation applied to Heschl's gyrus modulates pitch discrimination. Front. Psychol. 1, 193. https://doi.org/10.3389/fpsyg.2010.00193.

Mitchel, A.D., Christiansen, M.H., Weiss, D.J., 2014. Multimodal integration in statistical learning: evidence from the McGurk illusion. Front. Psychol. 5, 407. https://doi.org/10.3389/fpsyg.2014.00407.

Mitchel, A.D., Weiss, D.J., 2011. Learning across senses: cross-modal effects in multisensory statistical learning. J. Exp. Psychol. Learn. Mem. Cognition 37 (5), 1081–1091. https://doi.org/10.1037/a0023700.

Molnar-Szakacs, I., Iacoboni, M., Koski, L., Mazziotta, J.C., 2005. Functional segregation within pars opercularis of the inferior frontal gyrus: evidence from fMRI studies of imitation and action observation. Cereb. Cortex 15 (7), 986–994. https://doi.org/10.1093/cercor/bhh199.

Näätänen, R., Pakarinen, S., Rinne, T., Takegata, R., 2004. The mismatch negativity (MMN): towards the optimal paradigm. Clin. Neurophysiology Official J. Int. Fed. Clin. Neurophysiology 115 (1), 140–144.

Nichols, E.S., Grahn, J.A., 2016. Neural correlates of audiovisual integration in music reading. Neuropsychologia 91, 199–210. https://doi.org/10.1016/j.neuropsychologia.2016.08.011.

Nieder, A., Miller, E., 2003. Coding of cognitive magnitude: compressed scaling of numerical information in the primate prefrontal cortex. Neuron 37 (1), 149–157.

Niso, G., Bruña, R., Pereda, E., Gutiérrez, R., Bajo, R., Maestú, F., del-Pozo, F., 2013. HERMES: towards an integrated toolbox to characterize functional and effective brain connectivity. Neuroinformatics 11 (4), 405–434. https://doi.org/10.1007/s12021-013-9186-1.

Oechslin, M.S., Imfeld, A., Loenneker, T., Meyer, M., Jäncke, L., 2010. The plasticity of the superior longitudinal fasciculus as a function of musical expertise: a diffusion tensor imaging study. Front. Hum. Neurosci. 3, 76. https://doi.org/10.3389/neuro.09.076.2009.

Oldfield, R., 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia Vol. 9. https://doi.org/10.1016/0028-3932(71)90067-4.

Pantev, C., Paraskevopoulos, E., Kuchenbuch, A., Lu, Y., Herholz, S.C., 2015. Musical expertise is related to neuroplastic changes of multisensory nature within the auditory cortex. Eur. J. Neurosci. 41 (October 2014), 709–717. https://doi.org/10.1111/ejn.12788.

Paraskevopoulos, E., Chalas, N., Bamidis, P.D., 2017. Functional connectivity of the cortical network supporting statistical learning in musicians and non-musicians: an MEG study. Sci. Rep. 7. https://doi.org/10.1038/s41598-017-16592-y.

Paraskevopoulos, E., Kuchenbuch, A., Herholz, S.C., Foroglou, N., Bamidis, P., Pantev, C., 2014a. Tones and numbers: a combined EEG-MEG study on the effects of musical expertise in magnitude comparisons of audiovisual stimuli. Hum. Brain Mapp. 35 (11), 5389–5400. https://doi.org/10.1002/hbm.22558.

Paraskevopoulos, E., Kuchenbuch, A., Herholz, S.C., Pantev, C., 2011. Statistical learning effects in musicians and non-musicians: an MEG study. Neuropsychologia 50 (2), 341–349. https://doi.org/10.1016/j.neuropsychologia.2011.12.007.

Paraskevopoulos, E., Kuchenbuch, A., Herholz, S.C., Pantev, C., 2012. Musical expertise induces audiovisual integration of abstract congruency rules. J. Neurosci. 32 (50), 18196–18203. https://doi.org/10.1523/jneurosci.1947-12.2012.

Paraskevopoulos, E., Kuchenbuch, A., Herholz, S.C., Pantev, C., 2014b. Multisensory integration during short-term music reading training enhances both uni- and multisensory cortical processing. J. Cognitive Neurosci. 26 (10), 2224–2238. https://doi.org/10.1162/jocn_a_00620.

Parise, C.V., Spence, C., 2009. "When birds of a feather flock together": synesthetic correspondences modulate audiovisual integration in non-synesthetes. PloS One 4 (5) e5664. https://doi.org/10.1371/journal.pone.0005664.

Pascual-Marqui, R., Michel, C., 1994. Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. Int. J. Psychophysiol. 18 (1), 49–65. https://doi.org/10.1016/0167-8760(84)90014-X.

Rodríguez-Fornells, A., Cunillera, T., Mestres-Missé, A., de Diego-Balaguer, R., 2009. Neurophysiological mechanisms involved in language learning in adults. Philosophical Trans. R. Soc. Lond. B Biol. Sci. 364, 1536.

Rosenblum, L.D., Dias, J.W., Dorsi, J., 2017. The supramodal brain: implications for auditory perception. J. Cognitive Psychol. 29 (1), 65–87. https://doi.org/10.1080/20445911.2016.1181691.

Saffran, J.R., Aslin, R.N., Newport, E.L., 1996. Statistical learning by 8-month-old infants. Science 274 (5294), 1926. https://doi.org/10.1126/science.274.5294.1926.

Saffran, J.R., Johnson, E.K., Aslin, R.N., Newport, E.L., 1999. Statistical learning of tone sequences by human infants and adults. Cognition 70 (1), 27–52. https://doi.org/10.1016/S0010-0277(98)00075-4.

Schapiro, A.C., Gregory, E., Landau, B., McCloskey, M., Turk-Browne, N.B., 2014. The necessity of the medial temporal lobe for statistical learning. J. Cognitive Neurosci. 26 (8), 1736–1747. https://doi.org/10.1162/jocn_a_00578.

Schlaug, G., Jäncke, L., Huang, Y., Staiger, J.F., Steinmetz, H., 1995. Increased corpus callosum size in musicians. Neuropsychologia 33 (8), 1047–1055. https://doi.org/10.1016/0028-3932(95)00045-5.

Schlaug, G., Norton, A., Overy, K., Winner, E., 2005. Effects of music training on the child's brain and cognitive development. Ann. N. Y. Acad. Sci. 1060, 219–230. https://doi.org/10.1196/annals.1360.015.

Schön, D., François, C., 2011. Musical expertise and statistical learning of musical and linguistic structures. Front. Psychol. 2, 167. https://doi.org/10.3389/fpsyg.2011.00167.

Shams, L., Seitz, A.R., 2008. Benefits of multisensory learning. Trends Cognitive Sci. 12 (11), 411–417. https://doi.org/10.1016/j.tics.2008.07.006.

Shams, L., Seitz, A., van Wassenhove, V., 2010. Audio-visual statistical learning. J. Vis. 6 (6), 152. https://doi.org/10.1167/6.6.152.

Stanford, T.R., Stein, B.E., 2007. Superadditivity in multisensory integration: putting the computation in context. NeuroReport 18 (8), 787–792. https://doi.org/10.1097/WNR.0b013e3280c1e315.

Steele, C.J., Bailey, J.A., Zatorre, R.J., Penhune, V.B., 2013. Early musical training and white-matter plasticity in the corpus callosum: evidence for a sensitive period. J. Neurosci. Official J. Soc. Neurosci. 33 (3), 1282–1290. https://doi.org/10.1523/JNEUROSCI.3578-12.2013.

Stekelenburg, J.J., Vroomen, J., 2007. Neural correlates of multisensory integration of ecologically valid audiovisual events. J. Cognitive Neurosci. 19 (12), 1964–1973. https://doi.org/10.1162/jocn.2007.19.12.1964.

Stekelenburg, J.J., Vroomen, J., 2015. Predictive coding of visual-auditory and motor-auditory events: an electrophysiological study. Brain Res. 1626, 88–96. https://doi.org/10.1016/j.brainres.2015.01.036.

Tanabe, H.C., Honda, M., Sadato, N., 2005. Functionally segregated neural substrates for arbitrary audiovisual paired-association learning. J. Neurosci. 25 (27), 6409–6418. https://doi.org/10.1523/JNEUROSCI.0636-05.2005.

Tesche, C., Uusitalo, M., Ilmoniemi, R., Huotilainen, M., Kajola, M., Salonen, O., Salonene, O., 1995. Signal-space projections of MEG data characterize both distributed and well-localized neuronal sources. Electroencephalogr. Clin. Neurophysiology 95 (3), 189–200. https://doi.org/10.1016/0013-4694(95)00064-6.

Tettamanti, M., Weniger, D., 2006. Broca's area: a supramodal hierarchical processor? Cortex 42 (4), 491–494. https://doi.org/10.1016/S0010-9452(08)70384-8.

Toiviainen, P., Tervaniemi, M., Louhivuori, J., Saher, M., Huotilainen, M., Näätänen, R., 1998. Timbre similarity: convergence of neural, behavioral, and computational approaches timbre similarity: convergence of neura and computational approaches. Music Percept. Interdiscip. J. 16 (2), 223–241.

Turk-Browne, N.B., Scholl, B.J., Chun, M.M., Johnson, M.K., 2009. Neural evidence of statistical learning: efficient detection of visual regularities without awareness. J. Cognitive Neurosci. 21 (10), 1934–1945. https://doi.org/10.1162/jocn.2009.21131.

Turk-Browne, N.B., Scholl, B.J., Johnson, M.K., Chun, M.M., 2010. Implicit perceptual anticipation triggered by statistical learning. J. Neurosci. Official J. Soc. Neurosci. 30 (33), 11177–11187. https://doi.org/10.1523/JNEUROSCI.0858-10.2010.

Uddén, J., Bahlmann, J., 2012. A rostro-caudal gradient of structured sequence processing in the left inferior frontal gyrus. Philosophical Trans. R. Soc. Lond. B Biol. Sci. 367 (1598).

Uno, T., Kawai, K., Sakai, K., Wakebe, T., Ibaraki, T., Kunii, N., Saito, N., 2015. Dissociated roles of the inferior frontal gyrus and superior temporal sulcus in audiovisual processing: top-down and bottom-up mismatch detection. PLoS ONE 10 (3) e0122580. https://doi.org/10.1371/journal.pone.0122580.

van Atteveldt, N., Formisano, E., Goebel, R., Blomert, L., 2004. Integration of letters and speech sounds in the human brain. Neuron 43 (2), 271–282. https://doi.org/10.1016/j.neuron.2004.06.025.

van Atteveldt, N.M., Formisano, E., Blomert, L., Goebel, R., 2006. The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. Cereb. Cortex 17 (4), 962–974. https://doi.org/10.1093/cercor/bhl007.

van Atteveldt, N., Murray, M.M., Thut, G., Schroeder, C.E., 2014. Multisensory integration: flexible use of general operations. Neuron 81 (6), 1240–1253. https://doi.org/10.1016/j.neuron.2014.02.044.

Warrier, C., Wong, P., Penhune, V., Zatorre, R., Parrish, T., Abrams, D., Kraus, N., 2009. Relating structure to function: Heschl's gyrus and acoustic processing. J. Neurosci. Offic. J. Soc. Neurosci. 29 (1), 61–69. https://doi.org/10.1523/JNEUROSCI.3489-08.2009.

Wolters, C.H., Anwander, A., Tricoche, X., Weinstein, D., Koch, M.A., MacLeod, R.S., 2006. Influence of tissue conductivity anisotropy on EEG/MEG field and return current computation in a realistic head model: a simulation and visualization study using high-resolution finite element modeling. NeuroImage 30 (3), 813–826. https://doi.org/10.1016/j.neuroimage.2005.10.014.

Zalesky, A., Fornito, A., Bullmore, E.T., 2010. Network-based statistic: identifying differences in brain networks. NeuroImage 53 (4), 1197–1207. https://doi.org/10.1016/j.neuroimage.2010.06.041.

Zatorre, R.J., Chen, J.L., Penhune, V.B., 2007. When the brain plays music: auditory-motor interactions in music perception and production. Nat. Rev. Neurosci. 8 (7), 547–558. https://doi.org/10.1038/nrn2152.