

Multisensory Integration during Short-term Music Reading Training Enhances Both Uni- and Multisensory Cortical Processing

Evangelos Paraskevopoulos¹, Anja Kuchenbuch¹, Sibylle C. Herholz²,
and Christo Pantev¹

Abstract

■ The human ability to integrate the input of several sensory systems is essential for building a meaningful interpretation out of the complexity of the environment. Training studies have shown that the involvement of multiple senses during training enhances neuroplasticity, but it is not clear to what extent integration of the senses during training is required for the observed effects. This study intended to elucidate the differential contributions of uni- and multisensory elements of music reading training in the resulting plasticity of abstract audiovisual incongruency identification. We used magnetoencephalography to measure the pre- and posttraining cortical responses of two randomly assigned groups of participants that followed either an audiovisual music reading training that required multisensory integration (AV-Int group) or a unisensory training that had separate auditory and

visual elements (AV-Sep group). Results revealed a network of frontal generators for the abstract audiovisual incongruency response, confirming previous findings, and indicated the central role of anterior prefrontal cortex in this process. Differential neuroplastic effects of the two types of training in frontal and temporal regions point to the crucial role of multisensory integration occurring during training. Moreover, a comparison of the posttraining cortical responses of both groups to a group of musicians that were tested using the same paradigm revealed that long-term music training leads to significantly greater responses than the short-term training of the AV-Int group in anterior prefrontal regions as well as to significantly greater responses than both short-term training protocols in the left superior temporal gyrus (STG). ■

INTRODUCTION

The human ability to integrate the input of several sensory systems is essential for building a meaningful interpretation out of the complexity of the environment and accurately adapting to it. The choice regarding which sensory information will be integrated in one percept may rely on physical characteristics of the stimuli, namely, temporal or spatial proximity (Lee & Noppeney, 2011; Innes-Brown & Crewther, 2009); it may be explicitly learned for specific stimuli (Naumer et al., 2009); or it may follow rules that bind the stimuli together (Besle, Hussain, Giard, & Bertrand, 2013). Studying the brain mechanisms underlying this integration is important for understanding perception within an ecologically valid framework.

Music notation reading provides a very useful model for studying multisensory phenomena, as it combines auditory and visual information in a highly structured manner (Stewart, 2005). A recent study by our group (Paraskevopoulos, Kuchenbuch, Herholz, & Pantev, 2012b) used magnetoencephalography (MEG) to identify the neural correlates of an audiovisual incongruency response generated be-

cause of the violation of abstract congruency rules. The chosen rule was comparable to musical reading: “the higher the pitch of the tone—the higher the position of the circle” and the incongruency response was mainly located in frontal regions. Plasticity effects because of long-term musical training on this response were investigated by comparing musicians to nonmusicians and were located in frontal, temporal, and occipital areas. A previous study by Paavilainen, Simola, Jaramillo, Näätänen, and Winkler (2001) regarding abstract conjunction of unisensory auditory features revealed that violation of the conjunction rule used (i.e., the higher the frequency, the louder the intensity of the tone) elicited an MMN response, an event-related component known to reflect discrimination of novel sound events from an expected input. Typically, MMN is generated in the auditory cortex around 120–250 msec after the onset of a deviant sound within a stream of standard sounds (Näätänen, 1995).

Nevertheless, a cross-sectional study comparing musicians to nonmusicians cannot be conclusive on the nature versus nurture debate, whereas a study design based on controlled training of novice participants with random group assignment allows causal inference on the plasticity effects (Lappe, Herholz, Trainor, & Pantev, 2008). More importantly, training studies offer the opportunity to

¹University of Münster, ²German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany

disentangle the elements of music training and investigate the contribution of each element in the resulting plasticity (Herholz & Zatorre, 2012). Therefore, directly comparing different types of training allows conclusions regarding task specificity of each training element and the resulting neuroplasticity.

Several recent studies argue that uni- and multisensory training have different effects in the resulting brain plasticity in structural and functional levels (Paraskevopoulos, Kuchenbuch, Herholz, & Pantev, 2012a; Butler, James, & James, 2011; Scholz, Klein, Behrens, & Johansen-Berg, 2009). Scholz et al. (2009) used a visuomotor task (juggling) to reveal that 6 weeks of multisensory training induces white matter changes that were not found in a group of controls. Nevertheless, this study compared the effects of visuomotor training to untrained controls, and therefore, conclusions regarding the crucial components of training that drive cortical plasticity cannot be drawn.

In a recent study, Butler et al. (2011) used fMRI to reveal that after short-term active learning that included auditory, visual, and motor information, the participants of the study had enhanced processing of both uni- and multisensory information as compared with a group that had passive audiovisual familiarization with the same objects. Moreover, Lappe et al. (Lappe, Trainor, Herholz, & Pantev, 2011; Lappe et al., 2008) used two training protocols based on playing (auditory sensorimotor training) or merely listening (auditory training) to the piano to show that the additional involvement of the sensorimotor system during the training resulted in greater neuroplastic changes than the unisensory training in MMN generation, as measured by MEG. Another recent training study by our group (Paraskevopoulos et al., 2012a) argued that plasticity because of short-term multisensory training alters the function of separate multisensory structures and not merely the unisensory ones along with their interconnection. In this study, musically naive participants were trained to play tone sequences from visually presented patterns in a music notation-like system (auditory-visual-somatosensory training), whereas another group received audiovisual training only that involved viewing the same patterns, attentively listening to the recordings of the first group sessions, and pressing the right or left foot pedal of the piano to indicate whether the heard tone sequences matched the visual patterns regarding the audiovisual binding rule (auditory-visual training). The cortical responses pre- and posttraining were assessed via MEG measurements of an audiovisual, an auditory, and a visual MMN response. The results of this study revealed enhancement of the audiovisual MMN although there was no significant effect in the auditory and visual mismatches. The studies of Paraskevopoulos et al. (2012b), Butler et al. (2011), and Lappe et al. (2008, 2011) allow identifying general and task-specific effects of multisensory training, in particular, the additive effects of different training components that engage the visual, auditory, and sensorimotor systems. Such comparisons of trainings that differ systematically

regarding training components are highly informative, as they allow attribution of effects to specific training components (Coffey & Herholz, 2013). However, because the focus in previous studies was on the additive effects of the sensorimotor component in the trainings, conclusions regarding the role of multisensory integration of auditory and visual information during training were not possible. This study closes this gap by specifically comparing combined audiovisual training with training that involved separate auditory and visual training components. Thus, rather than adding another modality to study its additive effect as in previous studies, the modalities involved in this study were the same across trainings, and the trainings differed regarding the amount of multisensory processing during training.

This study intended to elucidate the differential contributions of the unisensory and the multisensory elements of music reading training in the resulting plasticity of the abstract audiovisual incongruency response presented in Paraskevopoulos et al. (2012b). Hence, our hypothesis was that uni- and multisensory training will result in different patterns of cortical plasticity. Specifically, we hypothesize that unisensory training will enhance the cortical responses generated from auditory and visual sources, whereas multisensory training will result in an enhancement of the responses generated in regions that are correlated with multisensory processing and have been found to contribute in the processing of audiovisual incongruencies, namely the prefrontal cortex. Two different short-term training protocols based on music reading were offered to two randomly assigned groups of non-musicians: concurrent audiovisual training, or separate auditory and visual training. The resulting plasticity was assessed via prior and post training MEG measurements of a multifeatured oddball paradigm that allowed, within one run, the evaluation of the responses in abstract audiovisual incongruencies, as well as auditory and visual mismatches. Additionally, the post training results of both groups were compared with the results of the group of musicians of our previous study (Paraskevopoulos et al., 2012b). This allowed us to compare the effects of experimentally controlled short-term training (present study) and of long-term training in reading music (musicians in the previous study) on the processing of abstract audiovisual incongruencies. This comparison was performed to test whether the effect of long-term musical training shows greater similarity to the concurrent audiovisual training or to the separate auditory and visual training.

METHODS

Participants

Thirty-three individuals participated in this study. Nine of them were excluded from further analysis because of extensive head movement during the first MEG measurement (head movement > 0.7 cm). The 24 remaining

individuals constituted the sample of the study (mean age = 26.45, $SD = 4.25$; seven men) and were equally and randomly divided into two groups: one following integrated audiovisual training (AV-Int) and the other separate auditory and visual training (AV-Sep). Participants had not received any musical education before their participation in the study, apart from the compulsory lessons in school. All participants were right-handed according to the Edinburgh Handedness Inventory (Oldfield, 1971) and had normal hearing as evaluated by clinical audiometry. Participants provided written consent before 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

MEG Recordings

The stimuli used in the MEG recordings were similar to the ones used in the previous study (Paraskevopoulos et al., 2012b). They consisted of different conditions (audiovisual congruent [standard], audiovisual incongruent, auditory deviant, and visual deviant). All conditions were prepared by combining five-tone melodies and five images representing the pitch height of each tone in a simplified music reading modus. The images representing the pitch height were constructed as follows: five white horizontal lines were presented against a black background (Figure 1) similar to the staff lines in music notation. One blue disk (RGB color codes: red: 86; green: 126; blue: 214) was then placed in the middle of the horizontal direction and in one of the four spaces between the lines. The five-tone melo-

dies were constructed by the combination of four sinusoidal tones (F5, 698.46 Hz; A5, 880.46 Hz; C6, 1046.50 Hz; and E6, 1318.51 Hz) with duration of 400 msec and 10 msec rise and decay time (48000 Hz, 16 bit). The ISI (stimulus offset to stimulus onset) between each tone was 500 msec, and the total duration of each melody was 4 sec. Eight different melodies were prepared for each condition. The first tone of all melodies was C5, and it was always congruent.

For the audiovisual congruent stimuli (standard), each tone of the melodies was combined with the corresponding image according to the rule “the higher the pitch, the higher the position of the disk,” similar to the Western music notation system. For each tone played, the corresponding image appeared at the same time and for the same duration. The audiovisual incongruent stimuli were prepared along the same principles as the congruent ones, except that one of the tones of the melodies was paired with an image that did not correspond to the tone according to the above-mentioned rule. The difference always violated the contour of the melody and not simply the interval. The auditory (timbre) deviant stimuli were also prepared along the same principles apart from the replacement of one of the tones with another of the same frequency produced with a sawtooth waveform and filtered with a low-pass filter at 5000 Hz. The visual (color) deviant stimuli had congruent audiovisual information, but the disk of one of the images was replaced with one of red color (RGB color codes: red: 214; green: 86; blue: 126).

Consequently in total there were four conditions of stimuli, all providing audiovisual stimulation: Condition 1 included no mismatches and therefore is referred as standard, Condition 2 included an audiovisual incongruency

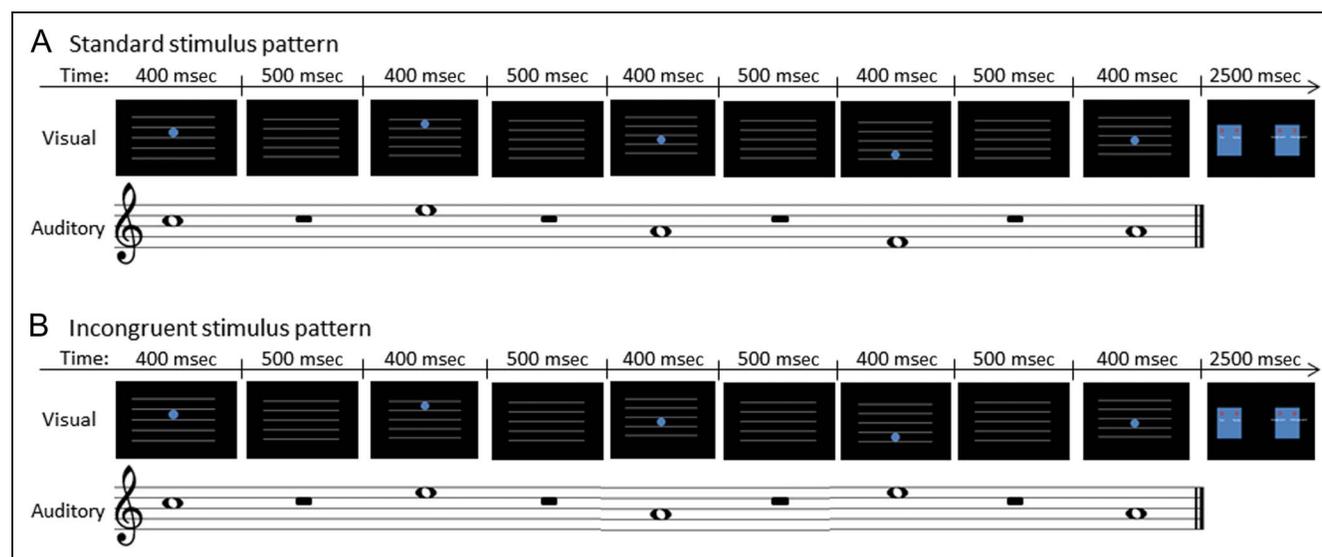


Figure 1. Example of an audiovisual congruent and an incongruent trial. (A) presents a congruent trial (that also served as auditory and visual standard). (B) presents an incongruent trial. The line “time” represents the duration of the presentation of the auditory and visual part of the stimulus. The last picture of each trial represents the intertrial stimulus in which participants had to answer if the trial was congruent or incongruent.

(violation of the rule: “the higher the pitch of the tone—the higher the position of the circle”), Condition 3 included an auditory mismatch (timbre), and Condition 4 included a visual mismatch (color).

Training

The stimuli used in the training protocol can be divided in three categories: (a) audiovisual (congruent and incongruent), used for the training of the AV-Int group; (b) auditory (standard and deviant); and (c) visual (standard and deviant); (b) and (c) were used for the training of the AV-Sep group. The audiovisual stimuli were constructed along the same principles of the audiovisual stimulus patterns used in the MEG recordings, but to avoid effects of training to the specific patterns the combinations of the five tones used for each pattern were new. The color of the presented disk also differed from the one used in the MEG recordings (red: 86, green: 126, blue: 214).

The auditory stimuli used as standards the same five-tone melodies that were used for the AV-Int group’s training stimuli but omitted the visual part of the stimulation. The auditory deviant tones were new. The deviancy consisted of differences in pitch height. The visual stimuli used as standards the corresponding part of the AV-Int group’s training stimuli, however, omitting the auditory stimulation. The visual deviant stimuli were again new. The deviancy consisted of differences in the height of the position of the disk, that is, instead of being between the score lines they were crossing the lines. Consequently, all incongruencies (and deviances) focused on the height of the pitch (auditory), the height of the position of the disk (visual), or their combination (audiovisual).

Each category of pattern had four different levels of difficulty based on a manipulation of the ISI and the saliency of incongruency (mismatch). For all categories of stimulation, the ISI used was 500 msec for the first level of training and 350 msec for the next ones. The saliency of the incongruency of the audiovisual stimuli was manipulated by altering the contour of the melodies in the first two levels of training (as in the MEG recordings) and the interval in the next ones (i.e., disk and tone were moving along the same direction but with different intervals). In the fourth level of difficulty, stimuli from all levels were randomly mixed. The saliency of the mismatch of the auditory stimuli was manipulated by using as deviant a pitch outside or inside of the range of the standard pitches (G6, 1567.98 Hz and B5, 932.33 Hz accordingly). The saliency of the mismatch of the visual stimuli was manipulated by changing the amount of area of the disc body crossing the score line. In the first two levels, 50% of the disk body area was crossing the score line, whereas in the other ones only 10%. As in the audiovisual training, in the fourth level of difficulty, stimuli from all levels were randomly mixed.

Experimental Design

MEG Recordings

Evoked magnetic fields were recorded with a 275-channel whole-head system (OMEGA, CTF Systems, Inc., Port Coquitlam, Canada) in a magnetically shielded room. Data were acquired continuously during each presentation block with a sampling rate of 600 Hz. Participants were seated upright, and their head position was comfortably stabilized with pads inside the MEG dewar. The auditory part of the stimuli was delivered via 60-cm-long silicon tubes at 60 dB SL above the individual hearing threshold that was determined with an accuracy of at least 5 dB at the beginning of each MEG session for each ear. The visual part of the stimuli was presented on a flat panel display (LG 1970 HR) located approximately 150 cm away from the participant’s nasion. The monitor was running at 60 Hz and with a spatial resolution of 1280 × 1024 pixels. The viewing angle ranged from -3.86° to 3.86° in the horizontal direction and from -1.15° to 1.15° in the vertical direction. The tones and disks were always presented synchronously, and the recording was synchronized to (a) the presentation of all tones of the audiovisual congruent patterns that served as standards for all modalities, (b) the incongruent tone of the audiovisual incongruent patterns serving as deviant in the audiovisual modality, (c) the deviant timbre tone of the auditory deviant patterns serving as deviants in the auditory modality, and (d) the deviant color disk of the visual deviant patterns, serving as deviants for the visual modality.

Both pre- and posttraining MEG recordings were identical. All four categories of stimulus patterns were randomly presented in one block. This block consisted of 26 presentations of each stimulus pattern category randomly interleaved, creating thus a multifaceted oddball paradigm (Nääänen, Pakarinen, Rinne, & Takegata, 2004) appropriately adapted for a multisensory experiment. This resulted in an incongruent (deviant) to congruent (standard) ratio of 20% for all modalities. Within 2.5 sec after each pattern presentation, participants had to answer via button pressings if the stimulus pattern presentation was congruent or incongruent (right hand) and if there was a tone sounding differently than all others or a disk of different color (left hand). Thus, there were two buttons per hand. During this intertrial interval, an image was presented to the participants reminding them which button represented each answer. Instructions for the task along with one example of each pattern category were given to the participants prior the beginning of the MEG recordings. Participants were exposed to four measurement blocks, lasting approximately 14.5 min each, with short breaks in between. The total number of stimulus patterns for each category within each block was 104.

Training

Five sessions of training took place during a period of 1 week. The first training was done directly after the

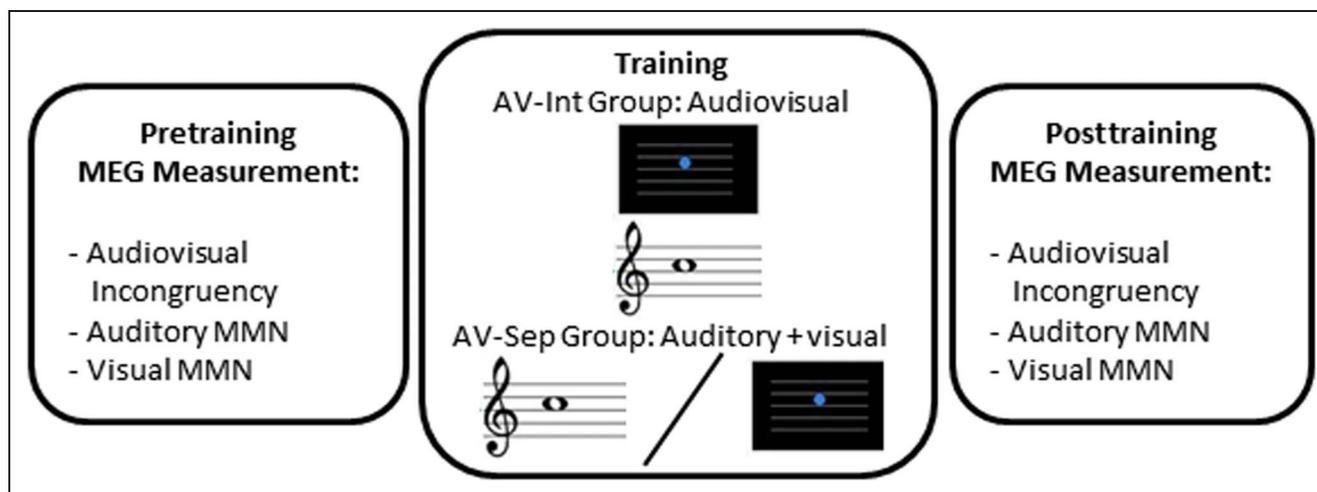


Figure 2. Illustration of the design of the study. An MEG measurement took place pre- and posttraining combining an audiovisual congruent, an audiovisual incongruent, an auditory deviant, and a visual deviant in a multifeature oddball paradigm. During the training, half of the participants received audiovisual training, and half auditory and visual.

pretraining MEG recording and the last immediately before the posttraining recording. Each training session lasted 28 min. The first three training sessions were divided in two levels of difficulty lasting 14 min each; the last two included only one level. Specifically, the first session included Levels 1 and 2, the second session Levels 2 and 3, the third one Levels 3 and 4, and the fourth and fifth only Level 4. Participants were seated in front of a computer while a screen adjusted approximately at the height of the nasion presented the stimulus patterns. After each pattern, participants of the AV-Int group had to judge if the audiovisual combination was congruent or incongruent by pressing the corresponding computer mouse button. Each training session of the AV-Sep group was further divided in two parts: one presenting the auditory stimulus patterns and the other the visual ones. The task of the AV-Sep group was to judge for each five-tone pattern (merely auditory) or five-disk pattern (merely visual) if it included only standard stimuli (tones or disks accordingly) or if a deviant was present. Responses of the AV-Sep group were also given via mouse buttons. Participants of both groups did not receive feedback regarding the accuracy of their responses (Figure 2).

Data Analysis

The BESA software (BESA research, version 5.3.7, Megis Software, Heidelberg, Germany) was used for initial processing of the MEG data. The recorded data were separated in epochs of 700 msec including a prestimulus interval of 200 msec. Epochs containing signals larger than 2.5 pT were considered artifact contaminated and excluded from the averaging. Data were filtered offline with a high-pass filter of 1 Hz, a low-pass filter of 30 Hz, and an additional notch filter at 50 Hz. Epochs were baseline corrected using the interval from -100 to 0 msec.

Averages of all four measurement blocks were computed separately for the standard and the incongruent stimuli of the audiovisual modality and for the deviants of the auditory and visual modalities.

Current density reconstructions (CDRs) were calculated on the neural responses of each participant for each stimulus category (congruent audiovisual, incongruent audiovisual, auditory deviant, visual deviant) using the LORETA method (Pascual-Marqui & Michel, 1994). LORETA directly computes a current distribution throughout the full brain volume instead of a limited number of dipolar point sources or a distribution restricted on the surface of the cortex. This method has been used successfully for the mapping of audiovisual incongruities (Paraskevopoulos et al., 2012a, 2012b) as well as unisensory MMN (Marco-Pallarés, Grau, & Ruffini, 2005; Waberski et al., 2001) and has the advantage of not needing an a priori definition of the number of activated sources. A time window of 40 msec was used for the CDR (130–170 msec). This time window was chosen on the basis of our previous study (Paraskevopoulos et al., 2012b) that presented this abstract audiovisual incongruency response. Additionally, this time window is often chosen in the MMN literature to reveal early responses showing deviance detection (Tse, Tien, & Penney, 2006; Rüsseler, Altenmüller, Nager, Kohlmetz, & Münte, 2001). Each individual's mean CDR image over the selected time window was calculated and projected onto a standard MRI template based on the Montreal Neurological Institute template. The images were smoothed, and their intensities were normalized by convolving an isotropic Gaussian kernel with 7 mm FWHM through BESA's smoothing utility.

The software Statistical Parametric Mapping 8 (SPM8, www.fil.ion.ucl.ac.uk/spm) and GLM-Flex (nmr.mgh.harvard.edu/harvardagingbrain/People/AaronSchultz/GLM_Flex.html) running on Matlab (MathWorks, Inc., Natick, MA) were used for the statistical analysis of the CDRs.

Specifically, using GLM-Flex a separate analysis was designed for each modality (audiovisual, auditory, and visual) to create a $2 \times 2 \times 2$ mixed model ANOVA with between-subject factor Group (AV-Int and AV-Sep) and within-subject factors Condition (standard and deviant) and Measurement (pre and post). We tested the main effect of Condition (mismatch response) and the Group \times Condition \times Measurement interaction. A Condition \times Measurement interaction within each group was calculated to investigate the effects of each kind of training. Results were constrained in gray matter using a mask, thereby keeping the search volume small and in physiologically reasonable areas. A permutation method for peak-cluster level error correction (AlphaSim) at $p < .05$ was applied for this whole head analysis, as implemented in REST software (Song et al., 2011), by taking into account the significance of the peak voxel (threshold $p < .005$ uncorrected) along with the cluster size (threshold size > 194 voxels), thereby controlling for multiple comparisons. The smoothness factor used for AlphaSim estimation was calculated from the corresponding residual image of each analysis. Visualization was done using MRIcron (www.mccauslandcenter.sc.edu/mricro/mricron/).

Moreover the posttraining results of both groups were compared with the group of musicians (15 participants) of our previous study (Paraskevopoulos et al., 2012b) that used the same paradigm to differentiate effects of long- and short-term training in the audiovisual processing of abstract rules. A 3×2 mixed model ANOVA with factors Group (musicians, AV-Int group, and AV-Sep group) and Condition (congruent and incongruent) was calculated using GLM-Flex. The main effect of Group and the Group \times Condition interaction was investigated to show if there are differences between the long-term trained musicians and the short-term trained participants of this study. The threshold was defined using AlphaSim peak-cluster level error correction at $p < .05$ by taking into account the significance of the peak voxel (threshold $p < .005$ uncorrected) along with the cluster size (threshold size > 212 voxels).

RESULTS

Behavioral Responses

Pretraining Testing

The results of the behavioral discrimination task in all conditions showed a high level of correct responses already in the pretraining testing. Specifically, the mean correct responses from all participants for the audiovisual modality were 92.417 ($SD = 11.71$), representing 88.86%; the mean correct responses for the auditory modality were 102.33 ($SD = 1.65$), representing 98.39%; and the mean correct responses for the visual modality were 101.35 ($SD = 4.87$), representing 97.45%. A comparison of the discriminability index d' of the pretraining responses of the two groups revealed no significant differences.

Pre- versus Posttraining Testing Comparison

The discriminability index d' was calculated for pre- and posttraining testing and entered in three 2×2 mixed model ANOVAs (one for each modality) with within-subject factor Testing (pre- and posttraining) and between-subject factor Group (AV-Int and AV-Sep), corrected for multiple comparisons using Bonferroni correction. The ANOVA results of the audiovisual modality showed a significant main effect of Testing, $F(1, 22) = 6.487$, $p < .05$, but no interaction of Group \times Testing, indicating that the training caused an equal increase for both groups in the ability to identify the audiovisual incongruities. The results of the ANOVAs for the auditory and the visual modality did not reach significance either for the main effect of Testing or for the interaction of Group \times Testing, indicating that the training did not further increase the ability of the participants to identify the auditory or visual deviant.

Training Performance

To investigate the training performance of the two groups, the discriminability index d' was calculated for each session. The results were entered in a 5×2 mixed model ANOVA with within-subject factor Session (Training Sessions 1–5) and between-subject factor Group (AV-Int and AV-Sep), corrected for multiple comparisons using Bonferroni correction. The results revealed a significant main effect of Session, $F(4, 88) = 7.034$, $p < .05$, but the interaction of Session \times Group was not significant, revealing that the two groups had equal performance increases throughout the training. A post hoc analysis of the performance on each session revealed that the participants' results on Session 2 were significantly lower from Session 4 (mean difference = $-.557$, $p > .05$) and Session 5 (mean difference = $-.582$, $p > .05$), and the results on Session 3 were significantly lower from Session 4 as well (mean difference = $-.520$, $p > .05$) and Session 5 (mean difference = $-.545$, $p > .001$). Jointly, these post hoc results indicate that the participants increased their performance in Sessions 4 and 5, which differed significantly from all other sessions, except from the first one.

MEG Results

Audiovisual Modality

The audiovisual incongruency response generators are shown by the results of the main effect of condition in the statistical analysis of the audiovisual modality. This analysis revealed a complex network of sources mainly located at frontal regions. Specifically, significant effects were found in a cluster located in ACC (peak coordinates right hemisphere: $x = 8$, $y = 24$, $z = 46$; $t(22) = 3.79$; cluster size = 387 voxels; $p < .05$ AlphaSim-corrected; peak coordinates left hemisphere: $x = -6$, $y = 30$,

$z = 28$; $t(22) = 3.45$; cluster size = 501 voxels; $p < .05$), a region in the left anterior prefrontal cortex (APC), specifically the superior frontal gyrus (peak coordinates: $x = -4$, $y = 68$, $z = 8$; $t(22) = 3.33$; cluster size = 1049 voxels; $p < .05$), one cluster in the left inferior frontal gyrus (peak coordinates: $x = -42$, $y = 32$, $z = 22$; $t(22) = 3.08$; cluster size = 229 voxels; $p < .05$) and one in the right STG (peak coordinates: $x = 49$, $y = 7$, $z = -10$; $t(22) = 3.89$; cluster size = 625 voxels; $p < .05$). These results are summarized in Table 1, and the statistical map is presented in Figure 3. All anatomical regions are defined using the AAL atlas (Tzourio-Mazoyer et al., 2002).

The statistical analysis of the audiovisual modality (Figure 3) also revealed a significant three-way interaction of Group \times Condition \times Measurement, thereby showing a differentiated effect of the trainings between the two groups. This effect was located in two clusters: one in the posterior part of the left superior frontal gyrus (peak coordinates: $x = 26$, $y = 56$, $z = 26$; $F(1, 22) = 12.35$; cluster size = 296 voxels; $p < .05$) and one in the left STG (peak coordinates: $x = -60$, $y = -46$, $z = 12$; $F(1, 22) = 10.64$; cluster size = 210 voxels; $p < .05$). These results are summarized in Table 2.

The subsequent analyses of Condition \times Measurement interaction within each group revealed that the AV-Int training significantly affected a cluster located in the posterior part of the left STG (peak coordinates: $x = -60$, $y = -46$, $z = 12$; $F(1, 22) = 11.67$; cluster size = 423 voxels; $p < .05$), and the AV-Sep training affected significantly two clusters: one located in the right frontal pole and specifically the middle frontal gyrus (peak coordinates: $x = 32$, $y = 52$, $z = 28$; $F(1, 22) = 11.44$; cluster

size = 236 voxels; $p < .05$) and one located in the right cingulate cortex (peak coordinates: $x = 16$, $y = 20$, $z = 68$; $F(1, 22) = 12.19$; cluster size = 280 voxels; $p < .05$). These results are summarized in Table 2.

The comparison of the posttraining results of the non-musicians (both groups separately) with the musicians group of the Paraskevopoulos et al. (2012b) study revealed a two-way interaction of Group \times Condition located in the left superior frontal gyrus (peak coordinates: $x = -25$, $y = 52$, $z = -6$; $F(1, 36) = 6.81$; cluster size = 372 voxels; $p < .05$). The post hoc t test comparison of the responses to the incongruent condition between the groups (using the same threshold) revealed that this interaction originated from a significantly greater activity of this region in the group of musicians in the incongruent condition when compared with the AV-Int group. This region revealed no significant effects in the post hoc t test that compared the group of musicians with the AV-Sep group. Moreover, the main effect of group in the comparison of musicians to the posttraining responses of nonmusicians revealed a significant activation located in the left STG (peak coordinates: $x = -24$, $y = 55$, $z = -1$; $F(1, 36) = 12.639$; cluster size = 658 voxels; $p < .05$) indicating that musicians showed increased activity in this region in response to the audiovisual stimuli independent of their congruency or incongruency. The post hoc t test comparison of the group effect between the musicians and each of the two other groups (again using the same threshold) revealed that this effect originated from a significantly higher activation of this region in the group of musicians compared with both other groups. The statistical maps of these results are presented in Figure 4.

Table 1. Generators of the Incongruency Response of the Audiovisual Modality and the MMN Responses of the Auditory and Visual Modalities

Modality	Location of Activation	Coordinates			Peak Voxel $t(22)$ Value	Cluster Size (No. of Voxels)
		x	y	z		
Audiovisual	Right superior temporal gyrus	49	7	-10	3.89	625
	Right cingulate gyrus	8	24	36	3.79	387
	Left cingulate gyrus	-6	30	28	3.45	501
	Left superior frontal gyrus	-4	68	8	3.33	1049
Auditory	Left inferior frontal gyrus	-42	32	22	3.08	229
	Left anterior cingulate gyrus	-6	20	28	6.81	1667
	Left superior frontal gyrus	-14	38	48	5.61	4437
	Right middle temporal gyrus	-52	-22	-12	5.43	2355
Visual	Left superior temporal gyrus	47	-4	-10	3.42	1067
	Middle cingulate cortex	0	-18	55	3.89	1318
	Left fusiform gyrus	-30	-66	-10	3.72	1338
	Right fusiform gyrus	44	-30	-28	3.68	810

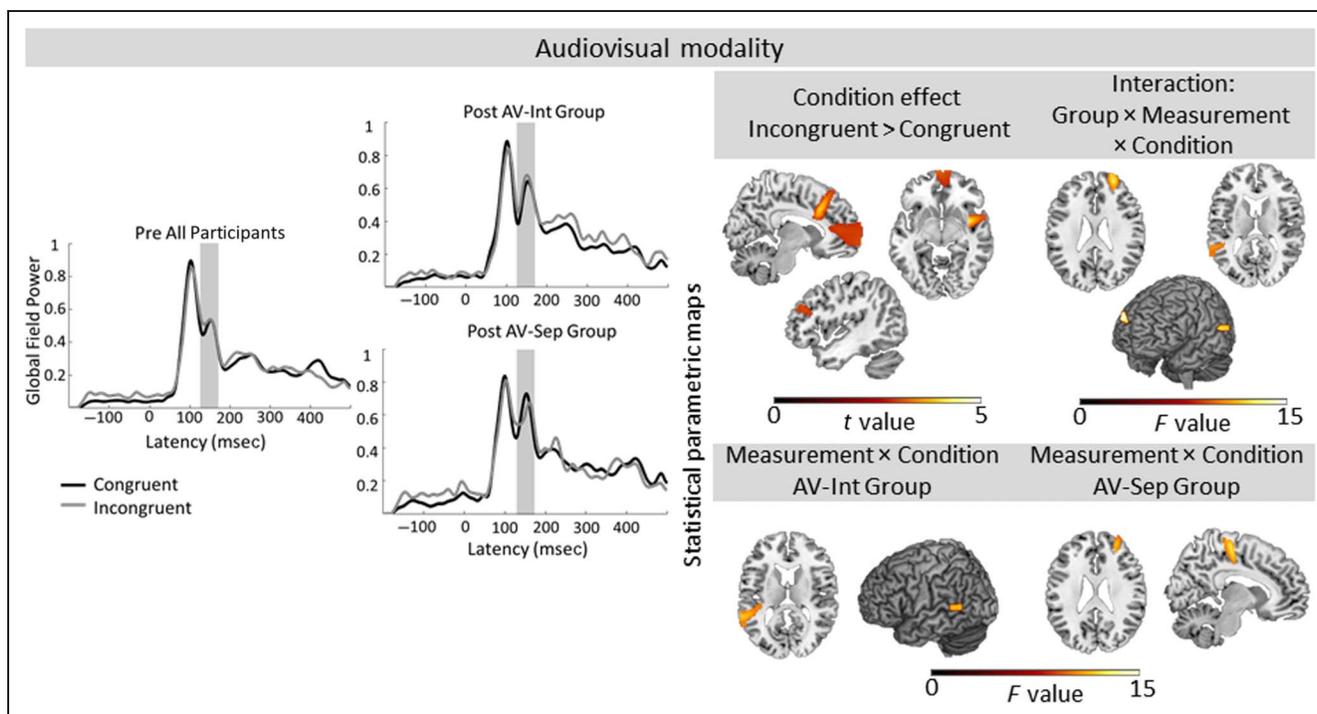


Figure 3. Top right: statistical parametric maps of the audiovisual incongruity response and the Group \times Measurement \times Condition interaction as revealed by the Flexible Factorial Model. Bottom right: statistical parametric maps of the training effects on the audiovisual incongruity response for each group. Threshold: AlphaSim corrected at $p < .05$ by taking into account peak voxel significance (threshold $p < .005$ uncorrected) along with cluster size (threshold size > 194 voxels). Left side: grand-averaged global field power for congruent (black line) and incongruent (gray line) response, pre (all participants) and post (divided by group) training. The time interval where the analysis was performed is marked gray.

Auditory Modality

The auditory MMN response generators are shown by the results of the main effect of Condition in the statistical analysis of the auditory modality. This analysis revealed significant effects in one cluster located in the right middle temporal gyrus (peak coordinates right side: $x = -52, y = -22, z = -12; t(22) = 5.43$; cluster size = 2355 voxels; $p < .05$), one in the left STG (peak coordinates: $x = -47, y = -4, z = -10; t(22) = 3.42$; cluster size = 1067 voxels; $p < .05$), a broad frontal region with a peak in the superior frontal gyrus (peak coordinates: $x = -14, y = 38, z = 48; t(22) = 5.61$; cluster size = 4437 voxels; $p < .05$), and the left anterior cingulate gyrus (peak coordinates: $x = -6, y = 20, z = 28; t(22) = 6.81$; cluster size = 1667 voxels; $p < .05$). These results are summarized in Table 1, and the statistical map is presented in Figure 5.

Furthermore, we found a significant three-way interaction of Group \times Condition \times Measurement, thereby showing a differentiated effect of the trainings between the two groups (Figure 5). These results were located in two clusters: one in the left STG (peak coordinates: $x = -58, y = -38, z = 10; F(1, 22) = 23.12$; cluster size = 1630 voxels; $p < .05$) and one in the left inferior temporal gyrus (peak coordinates: $x = -54, y = -18, z = -32; F(1, 22) = 16.43$; cluster size = 528 voxels; $p < .05$). These results are summarized in Table 2.

The subsequent analyses of Condition \times Measurement interaction within each group revealed that the AV-Int training significantly affected a cluster located in the posterior part of the left STG (peak coordinates: $x = -60, y = -34, z = 10; F(1, 22) = 12.37$; cluster size = 208 voxels; $p < .05$ AlphaSim corrected), and the AV-Sep training significantly affected three regions: One cluster was located in the left STG (peak coordinates: $x = -38, y = -38, z = 12; F(1, 22) = 17.95$; cluster size = 1398 voxels; $p < .05$), the second was located in the inferior temporal gyrus (peak coordinates: $x = -50, y = -14, z = -34; F(1, 22) = 14.75$; cluster size = 1061 voxels; $p < .05$), and the third was located in the right inferior frontal gyrus (peak coordinates: $x = 32, y = 20, z = -2; F(1, 22) = 16.94$; cluster size = 461 voxels; $p < .05$). These results are summarized in Table 2.

Visual Modality

The main effect of Condition of the statistical analysis for the visual MMN response revealed three clusters of activity. Two of the clusters were located bilaterally in the left and right fusiform gyrus (peak coordinates for the left hemisphere: $x = -30, y = -66, z = -10; t(22) = 3.72$; cluster size = 1338 voxels; $p < .05$; and peak coordinates for the right hemisphere: $x = 44, y = -30, z = -28; t(22) = 3.68$; cluster size = 810 voxels; $p < .05$) and

Table 2. Locations of Activities in Group \times Measurement \times Condition Interaction for Each Modality

Interaction Modality	Location of Activation	Coordinates			Peak Voxel $F(1, 22)$ Value	Cluster Size (No. of Voxels)
		x	y	z		
<i>Group \times Measurement \times Condition Interaction</i>						
Audiovisual	Left superior frontal gyrus	26	56	26	12.35	296
	Left superior temporal gyrus	-60	-46	12	10.64	210
Auditory	Left superior temporal gyrus	-58	-38	10	23.12	1630
	Left inferior temporal gyrus	-54	-18	-32	16.43	528
<i>Measurement \times Condition Interaction: AV-Int Group</i>						
Audiovisual	Left superior temporal gyrus	-60	-46	12	11.67	423
Auditory	Left superior temporal gyrus	-60	-34	10	12.37	208
<i>Measurement \times Condition Interaction: AV-Sep Group</i>						
Audiovisual	Right middle frontal gyrus	32	52	28	11.44	236
	Cingulate gyrus	16	20	68	12.19	280
Auditory	Left superior temporal gyrus	-38	-38	12	17.95	1398
	Left inferior temporal	-50	-14	-34	14.75	1061
	Right inferior frontal	32	20	-2	16.94	461
<i>Measurement \times Condition Interaction: Complete Sample</i>						
Visual	Left superior temporal gyrus	-60	-34	10	10.64	210

one in the interhemispheric fissure, middle cingulate cortex (peak coordinates: $x = -4, y = -22, z = 70; t(28) = 4.52$; cluster size = 1318 voxels; $p < .05$). These results are summarized in Table 1, and the statistical map is presented in Figure 6.

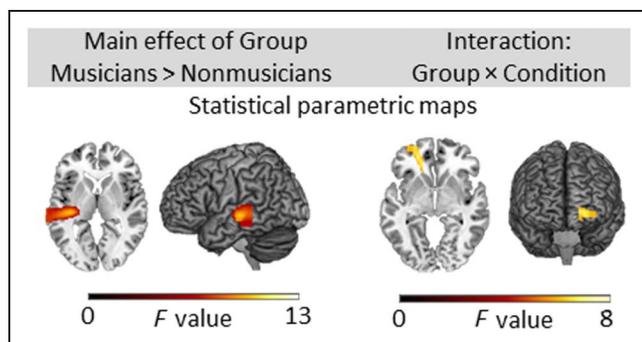


Figure 4. Statistical parametric maps of the comparison between musicians and the posttraining results of both groups of nonmusicians (AV-Int and AV-Sep) in the audiovisual incongruency response. The main effect of group and the interaction of Group \times Condition are presented. Threshold: AlphaSim-corrected at $p < .05$ by taking into account peak voxel significance (threshold $p < .005$ uncorrected) along with cluster size (threshold size > 212 voxels).

The statistical analysis of the visual MMN did not show significant results on the three-way interaction of Group \times Condition \times Measurement, indicating thus that the two kinds of training had no difference in the way they affected the identification of the visual deviants. Therefore, subsequently instead of analyzing the effect of training within each group, the interaction of Condition \times Measurement within the complete sample was analyzed. This analysis revealed that in both groups the training affected one region in the right frontal pole, specifically, the right middle frontal gyrus (peak coordinates: $x = 34, y = 54, z = 4; F(1, 22) = 19.55$; cluster size = 1297 voxels; $p < .05$). These results are summarized in Table 2.

DISCUSSION

This study intended to differentiate the contribution of uni- and multisensory elements of music reading training to audiovisual processing. We assessed the differences in training-related cortical plasticity in the processing of abstract audiovisual incongruencies. We used MEG to measure the cortical responses pre- and posttraining of two groups of participants that followed either a multisensory audiovisual music reading training (AV-Int group) or a protocol that contained separate auditory and visual

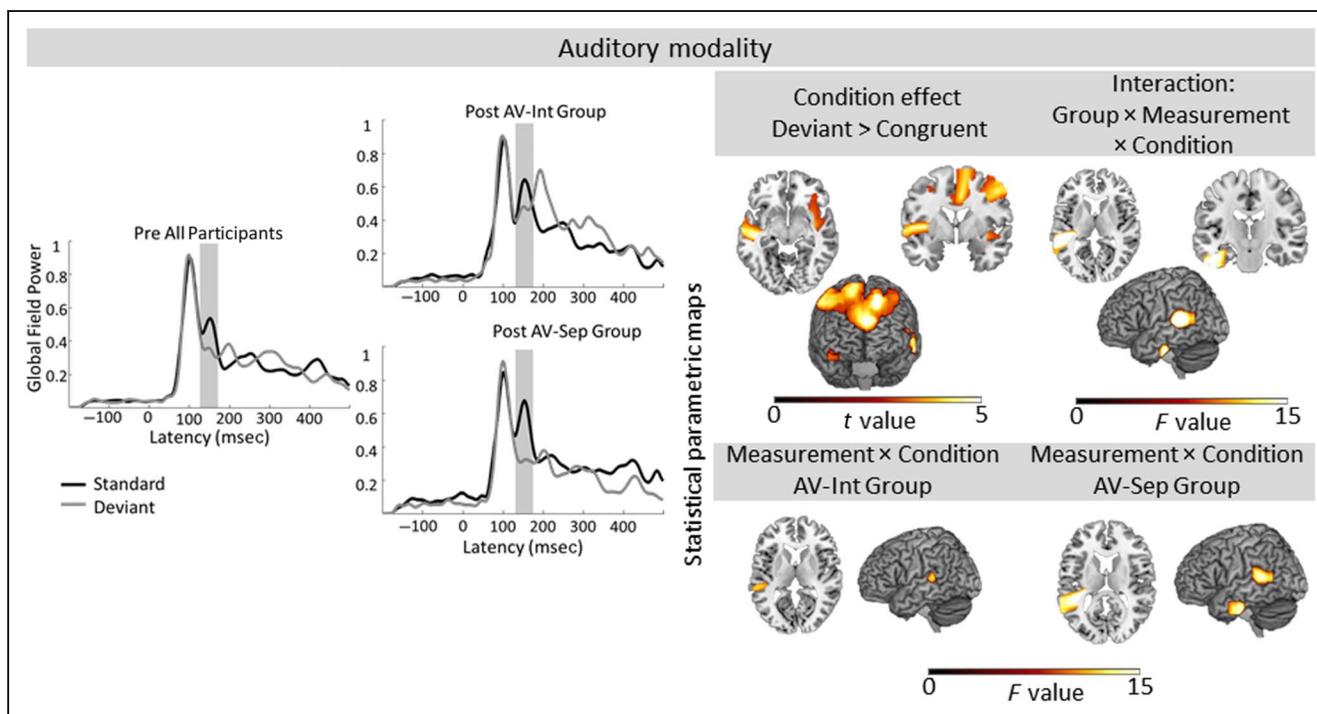


Figure 5. Top right: statistical parametric maps of the auditory mismatch and the Group \times Measurement \times Condition interaction as revealed by the Flexible Factorial Model. Bottom right: statistical parametric maps of the training effects on the auditory mismatch response for each group. Threshold: AlphaSim-corrected at $p < .05$ by taking into account peak voxel significance (threshold $p < .005$ uncorrected) along with cluster size (threshold size > 194 voxels). Left side: grand-averaged global field power for standard (black line) and deviant (gray line) response, pre (all participants) and post (divided by group) training. The time interval where the analysis was performed is marked gray.

training sections (AV-Sep group). In both types of training, discrimination abilities in both sensory modalities were equally trained, but audiovisual integration during training was only required in the AV-Int group.

Our results not only confirmed the neural generators of the abstract audiovisual incongruity response (Paraskevopoulos et al., 2012b) but, as a new and original result, also showed significant differences in this response

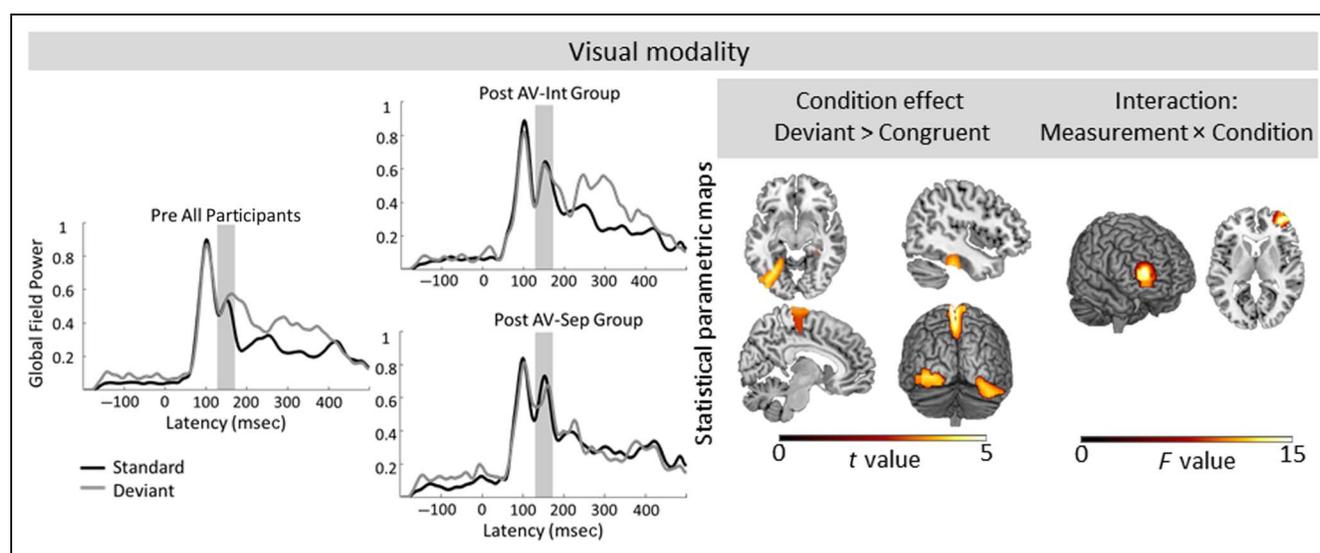


Figure 6. Right: statistical parametric maps of the visual mismatch response and the Measurement \times Condition interaction (training effect) as revealed by the Flexible Factorial Model. Threshold: AlphaSim-corrected at $p < .05$ by taking into account peak voxel significance (threshold $p < .005$ uncorrected) along with cluster size (threshold size > 194 voxels). Left: grand-averaged global field power for standard (black line) and deviant (gray line) response, pre (all participants) and post (divided by group) training. The time interval where the analysis was performed is marked gray.

between the two types of training in frontal and temporal regions. Moreover, a comparison of the posttraining cortical responses of both groups to a group of musicians that were tested using the same paradigm revealed that long-term music training leads to significantly greater responses than the short-term training of the AV-Int group in anterior prefrontal regions, as well as to significantly greater responses than both short-term training protocols in the left STG.

Training Effects on Behavioral Measures of Audiovisual Integration

The results of the behavioral measurements before the training were significantly higher than the chance level, confirming thus that the task was appropriate for a sample of nonmusicians. This finding was also present in our previous study (Paraskevopoulos et al., 2012b), and it has been additionally demonstrated in psychophysical studies of audiovisual correspondences between the height of a pitch and the height of a visual stimulus (Spence, 2011; Evans & Treisman, 2010; Walker et al., 2010). A recent psychophysical study (Chiou & Rich, 2012) demonstrated that high and low tones may induce an attention shift to upper or lower locations of the visual field and that this pitch-induced cuing effect is susceptible to contextual manipulations and volitional control. Additionally, our results are in line with a recent study indicating that when indirect tasks are used to measure this existing correspondence, musicians are superior to nonmusicians (Rusconi, Kwan, Giordano, Umiltà, & Butterworth, 2006).

Both types of training resulted in increased accuracy in the posttraining performance compared with pretraining, but only for the audiovisual modality, not for auditory and visual mismatches. This finding was expected because of the focus of the training in the height of the pitch, the height of the visually presented disk, and their combination. The identification of color and timbre was not explicitly trained. Additionally, the behavioral results of the auditory and visual modality were of high accuracy already in the pretraining measurement, not allowing thus easily further improvement and showing a ceiling effect.

Neural Correlates of Audiovisual, Auditory, and Visual Processing

The generators of the abstract audiovisual incongruency response, as revealed by the MEG results, confirm the main findings of our previous study (Paraskevopoulos et al., 2012b), indicating thus an important stability of the responses across different samples of participants. Specifically, the generators of this response consist of a frontal network of sources including the ACC, the left inferior frontal gyrus, the APC, and the right STG. The ACC has been shown previously to be directly connected

with audiovisual integration (Benoit, Raji, Lin, Jääskeläinen, & Stufflebeam, 2010) as well as attention load (Bush, Luu, & Posner, 2000; Pardo & Fox, 1991). The left inferior frontal gyrus is correlated with many different processes including working memory (Courtney & Ungerleider, 1997) and music reading (Stewart et al., 2003). The APC (Brodmann's area 10) is a region that has direct connections with STG and cingulate cortex, as a recent diffusion tensor imaging study revealed (Liu et al., 2013), and its functional role is proposed to be the maintenance of a goal while exploring and processing secondary goals (Koechlin, Basso, Pietrini, Panzer, & Grafman, 1999) or the integration of the results of two or more separate cognitive operations that are required to fulfill a behavioral goal (Ramnani & Owen, 2004). This interpretation would give to this region a central role in the process of identifying abstract audiovisual incongruencies. The right STG activation was not present in our previous study (Paraskevopoulos et al., 2012b). Nevertheless, STG is a region that is well known for responding to audiovisual stimuli (Werner & Noppeney, 2010; Ghazanfar & Schroeder, 2006), and it has been correlated with the identification of audiovisual mismatches (Paraskevopoulos et al., 2012a; Yumoto et al., 2005). It must be noted here that our paradigm requires that the auditory and visual stimulus patterns are jointly processed, but it does not conclusively demonstrate multisensory integration on a neural level. In two recent studies, Foster and Zatorre (2010a, 2010b) revealed that the intraparietal sulcus along with the right auditory cortex play a crucial role in transforming musical pitch information. However, we did not find these two regions (i.e., right auditory cortex and intraparietal sulcus) to be differently activated in the standard and incongruent stimulus patterns of this study. We hypothesize that the cognitive process of relative pitch comparisons may have been engaged in both standard and incongruent conditions to a similar extent.

The auditory mismatch condition showed increased activity bilaterally in STG, which is typically associated with auditory MMN (Alho, Woods, Algazi, Knight, & Näätänen, 1994; Näätänen, Gaillard, & Mäntysalo, 1978), along with a dorsolateral prefrontal activation that can be interpreted as reflecting attention switching between the multisensory condition (audiovisual) of our task and the unisensory one (auditory MMN; Paraskevopoulos et al., 2012b). The location of the activity in the visual MMN response was found bilaterally in the fusiform gyrus along with the middle cingulate cortex (for a recent review on visual MMN, see Kimura, Schröger, & Czigler, 2011). This activation pattern for the visual MMN elicited by color deviants is in line with the role of fusiform gyrus for color perception (Hsu, Kraemer, Oliver, Schlichting, & Thompson-Schill, 2011) and for visual working memory (Ungerleider, Courtney, & Haxby, 1998). A recent MEG study (Urakawa, Inui, Yamashiro, & Kakigi, 2010) of visual MMN based on color deviance localized the neural generators of this response in the middle occipital gyrus. The difference in the localization of the visual MMN in the two studies may be because

of the difference between the corresponding paradigms. The paradigm used in the study of Urakawa et al. (2010) induced a change in the periphery of the visual field by alternating the color of a series of LEDs adjacent to a screen that presented a movie. In our paradigm, the stimuli were in the center of the visual field, and the deviants and standards were the sole visual stimuli.

Training Effects on Neuronal Correlates of Audiovisual Integration, Auditory, and Visual Processing

The short-term training effects of the audiovisual modality reveal interesting interactions between the two types of training. Specifically, it seems that the audiovisual training resulted in an increase of the activity in the posterior STG, a region that is known to respond to concurrent audiovisual stimulation (Butler et al., 2011; Plank, Rosengarth, Song, Ellermeier, & Greenlee, 2011; Driver & Noesselt, 2008; Barraclough, Xiao, Baker, Oram, & Perrett, 2005) and is affected by musical training (Zimmerman & Lahav, 2012). On the contrary, the auditory and visual training showed an increase in the activity of the right APC a region, which, as already noted, is related to the coordination of two or more separate cognitive operations (Ramnani & Owen, 2004). Consequently, the differential results of audiovisual versus auditory and visual training seem to be very well suited to the corresponding demands of each training, that is, audiovisual integration for the AV-Int group whereas coordination integration of auditory and visual input in the AV-Sep group. The AV-Sep group additionally showed an increase in the cingulate gyrus, a region related with increased attention demands (Liu, Banich, Jacobson, & Tanabe, 2004). Nevertheless, the two types of training did not differ significantly with regard to the cingulate gyrus effects according to the results of the interaction of Group \times Condition \times Measurement.

The three-way interaction of Group \times Condition \times Measurement of the auditory modality revealed that the two types of training affected differently two cortical regions: one relatively broad region in the left posterior STG and one in the left inferior temporal gyrus. The within-group analysis of the auditory modality indicated that the concurrent audiovisual training affected the posterior STG, the same region that was shown to be affected in this group with regard to the audiovisual modality. On the contrary, the group that was trained with auditory and visual tasks showed a pattern of training effects that differed from the ones that the same group revealed in the training effects of the audiovisual task of the paradigm. Specifically, whereas an effect of training was present in the AV-Sep group in the APC and the cingulate gyrus when comparing the pre- and postmeasurement of the audiovisual incongruency response, a relatively broad region in the posterior part of the left STG along with a region in the left inferior temporal gyrus showed

significant differences between the pre- and posttraining measurements of the auditory mismatch response of the AV-Sep group. The increase of activation in the left STG is related with training in pitch identification (Zatorre & Gandour, 2008; Gaab, Gaser, & Schlaug, 2006), and the gray matter volume of this region is positively correlated with musicianship (Gaser & Schlaug, 2003). The latter is true also for inferior temporal gyrus (Gaser & Schlaug, 2003), a region that includes anatomical structures involved in the ventral visual stream (Cohen & Dehaene, 2004), shows increased activity when participants learn to choose actions prompted by visual stimuli (Passingham & Toni, 2001) and is correlated with visual long-term memory for spatial location (Moscovitch, Kapur, Köhler, & Houle, 1995). The two within-group analyses revealed that the differences found in the three-way interaction originate from a stronger training effect in the AV-Sep group when compared with the AV-Int group. The *F* values of this comparison in the AV-Sep group are stronger than the corresponding *F* values of the AV-Int group and cover a broader region of the posterior STG, whereas additionally there is a significant effect in the inferior temporal gyrus in the AV-Sep group that is not present in the AV-Int group.

The results of the training effects in the visual modality do not differ between the groups. Nevertheless, a significant effect of training was revealed in the right APC, being common for both groups. As mentioned above, this region is correlated with the coordination of different cognitive operations (Ramnani & Owen, 2004), a process that constitutes a suitable interpretation of the demands of the applied paradigm. The absence of a difference between the two groups can be attributed to the form of the visual training, which was probably not as demanding as the auditory or the audiovisual one, because the identification of spatial differences in the height of objects such as the ones used in the training stimuli is apparently an easy task for the visual system.

Short- versus Long-term Training Effects

The comparison of the posttraining responses of both groups (AV-Int and AV-Sep) to a group of musicians revealed that musicians' long-term training resulted in a significantly increased activation of the left APC when identifying the abstract audiovisual incongruencies in comparison with the group that received the concurrent audiovisual training (AV-Int). On the contrary, no difference was evident when compared with the group that received auditory and visual training (AV-Sep). Both the location and the group differences of this result are interesting. This region constitutes the left homologue of the region that was affected by the training of the AV-Sep group. Therefore, it seems reasonable to assume that the characteristics of the musical training correspond better to the characteristics of the training that the AV-Sep group received. Indeed, when reading a musical

score, musicians do not receive concurrent audiovisual stimulation. On the basis of the visual input from the musical score, musicians form a motor plan (along with an auditory expectation) that when executed will produce the actual corresponding auditory input. This auditory input then functions as a feedback regarding the accuracy of the performance and is compared with the initial visual input. The delay between the visual and the auditory input probably results in a greater need for coordinating and keeping active different cognitive processes and thus training the APC.

Moreover, the main effect of Group in this comparison revealed that musicians activated the left STG in greater extent than both short-term trained groups, independently of the congruency (or incongruency) of the stimulation, thus showing a more generalized effect of long-term musical training. This result corresponds well to anatomical studies that show increased gray matter volume in the left STG of musicians compared with controls (Hyde et al., 2009; Gaser & Schlaug, 2003).

Conclusion

The findings of this study reveal an important stability of the frontal network of sources that generate this abstract audiovisual incongruency response across different samples and indicate a central role for the contribution of APC in this process. The training effects indicated that even short-term training in music reading results in plastic changes in this region. Moreover, the training effects in the posterior part of the STG show a task-training preference by indicating a stronger contribution of the concurrent audiovisual training in the audiovisual task of the paradigm and a greater effect of the separate auditory and visual training in the auditory task. Thereby, the results of this study go beyond previous findings by conclusively elucidating the differential contribution of uni- and multi-sensory elements of music reading training in the resulting plasticity.

The comparison of a group of musicians to the post-training results of both groups used in this study revealed significant differences in the process of abstract audiovisual incongruency identification in left superior temporal and anterior prefrontal regions, but only when compared with the group that underwent the concurrent audiovisual training. This led to the interpretation that the delay between the visual and the auditory input that is present in music training when reading and playing a musical score results in a greater need for coordinating and keeping active different cognitive processes and thus training the APC.

Acknowledgments

We would like to thank our test participants for their cooperation and our technicians for supporting the data acquisition. This work was supported by the Deutsche Forschungsgemeinschaft

[PA392/12-2 and HE6067-1/1]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Reprint requests should be sent to Univ.-Prof. Dr. Christo Pantev, Institute for Biomagnetism and Biosignalanalysis, University of Münster, Malmedyweg 15, D-48149 Münster, Germany, or via e-mail: pantev@uni-muenster.de.

REFERENCES

- Alho, K., Woods, D., Algazi, A., Knight, R., & Näätänen, R. (1994). Lesions of frontal cortex diminish the auditory mismatch negativity. *Electroencephalography*, *91*, 353–362.
- Barracough, N. E., Xiao, D., Baker, C. I., Oram, M. W., & Perrett, D. I. (2005). Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions. *Journal of Cognitive Neuroscience*, *17*, 377–391.
- Benoit, M., Raji, T., Lin, F., Jääskeläinen, I. P., & Stufflebeam, S. (2010). Primary and multisensory cortical activity is correlated with audiovisual percepts. *Human Brain Mapping*, *31*, 526–538.
- Besle, J., Hussain, Z., Giard, M., & Bertrand, O. (2013). The representation of audiovisual regularities in the human brain. *Journal of Cognitive Neuroscience*, *25*, 365–373.
- Bush, G., Luu, P., & Posner, M. I. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends in Cognitive Sciences*, *4*, 215–222.
- Butler, A. J., James, T. W., & James, K. H. (2011). Enhanced multisensory integration and motor reactivation after active motor learning of audiovisual associations. *Journal of Cognitive Neuroscience*, *23*, 3515–3528.
- Chiou, R., & Rich, A. N. (2012). Cross-modality correspondence between pitch and spatial location modulates attentional orienting. *Perception*, *41*, 339–353.
- Coffey, E. B. J., & Herholz, S. C. (2013). Task decomposition: A framework for comparing diverse training models in human brain plasticity studies. *Frontiers in Human Neuroscience*, *7*, 640.
- Cohen, L., & Dehaene, S. (2004). Specialization within the ventral stream: The case for the visual word form area. *Neuroimage*, *22*, 466–476.
- Courtney, S., & Ungerleider, L. (1997). Transient and sustained activity in a distributed neural system for human working memory. *Nature*, *386*, 608–611.
- Driver, J., & Noesselt, T. (2008). Multisensory interplay reveals crossmodal influences on “sensory-specific” brain regions, neural responses, and judgments. *Neuron*, *57*, 11–23.
- Evans, A. C., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision*, *10*, 1–12.
- Foster, N. E. V., & Zatorre, R. J. (2010a). A role for the intraparietal sulcus in transforming musical pitch information. *Cerebral Cortex*, *20*, 1350–1359.
- Foster, N. E. V., & Zatorre, R. J. (2010b). Cortical structure predicts success in performing musical transformation judgments. *Neuroimage*, *53*, 26–36.
- Gaab, N., Gaser, C., & Schlaug, G. (2006). Improvement-related functional plasticity following pitch memory training. *Neuroimage*, *31*, 255–263.
- Gaser, C., & Schlaug, G. (2003). Brain structures differ between musicians and nonmusicians. *The Journal of Neuroscience*, *23*, 9240–9245.

- Ghazanfar, A. A., & Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends in Cognitive Sciences*, *10*, 278–285.
- Herholz, S. C., & Zatorre, R. J. (2012). Musical training as a framework for brain plasticity: Behavior, function, and structure. *Neuron*, *76*, 486–502.
- Hsu, N. S., Kraemer, D. J. M., Oliver, R. T., Schlichting, M. L., & Thompson-Schill, S. L. (2011). Color, context, and cognitive style: Variations in color knowledge retrieval as a function of task and subject variables. *Journal of Cognitive Neuroscience*, *23*, 2544–2557.
- Hyde, K. L., Lerch, J., Norton, A., Forgeard, M., Winner, E., Evans, A. C., et al. (2009). Musical training shapes structural brain development. *The Journal of Neuroscience*, *29*, 3019–3025.
- Innes-Brown, H., & Crewther, D. (2009). The impact of spatial incongruence on an auditory-visual illusion. *PLoS One*, *4*, e6450.
- Kimura, M., Schröger, E., & Czigler, I. (2011). Visual mismatch negativity and its importance in visual cognitive sciences. *NeuroReport*, *22*, 669–673.
- Koechlin, E., Basso, G., Pietrini, P., Panzer, S., & Grafman, J. (1999). The role of the anterior prefrontal cortex in human cognition. *Nature*, *399*, 148–151.
- Lappe, C., Herholz, S. C., Trainor, L. J., & Pantev, C. (2008). Cortical plasticity induced by short-term unimodal and multimodal musical training. *The Journal of Neuroscience*, *28*, 9632–9639.
- Lappe, C., Trainor, L. J., Herholz, S. C., & Pantev, C. (2011). Cortical plasticity induced by short-term multimodal musical rhythm training. *PLoS One*, *6*, e21493.
- Lee, H. L., & Noppeney, U. (2011). Long-term music training tunes how the brain temporally binds signals from multiple senses. *Proceedings of the National Academy of Sciences, U.S.A.*, *108*, E1441–E1450.
- Liu, H., Qin, W., Li, W., Fan, L., Wang, J., Jiang, T., et al. (2013). Connectivity-based parcellation of the human frontal pole with diffusion tensor imaging. *Journal of Neuroscience*, *33*, 6782–6790.
- Liu, X., Banich, M., Jacobson, B., & Tanabe, J. (2004). Common and distinct neural substrates of attentional control in an integrated Simon and spatial Stroop task as assessed by event-related fMRI. *NeuroImage*, *22*, 1097–1106.
- Marco-Pallarés, J., Grau, C., & Ruffini, G. (2005). Combined ICA-LORETA analysis of mismatch negativity. *NeuroImage*, *25*, 471–477.
- Moscovitch, C., Kapur, S., Köhler, S., & Houle, K. (1995). Distinct neural correlates of visual long-term memory for spatial location and object identity: A positron emission tomography study in humans. *Proceedings of the National Academy of Sciences, U.S.A.*, *92*, 3721–3725.
- Näätänen, R. (1995). The mismatch negativity: A powerful tool for cognitive neuroscience. *Ear and Hearing*, *16*, 6.
- Näätänen, R., Gaillard, A. W. K., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica*, *42*, 313–329.
- Näätänen, R., Pakarinen, S., Rinne, T., & Takegata, R. (2004). The mismatch negativity (MMN): Towards the optimal paradigm. *Clinical Neurophysiology*, *115*, 140–144.
- Naumer, M. J., Doehrmann, O., Müller, N. G., Muckli, L., Kaiser, J., & Hein, G. (2009). Cortical plasticity of audiovisual object representations. *Cerebral Cortex*, *19*, 1641–1653.
- Oldfield, R. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*, 97–113.
- Paavilainen, P., Simola, J., Jaramillo, M., Näätänen, R., & Winkler, I. (2001). Pre-attentive extraction of abstract feature conjunctions from auditory stimulation as reflected by the mismatch negativity (MMN). *Psychophysiology*, *38*, 359–365.
- Paraskevopoulos, E., Kuchenbuch, A., Herholz, S. C., & Pantev, C. (2012a). Evidence for training-induced plasticity in multisensory brain structures: An MEG study. *PLoS One*, *7*, e36534.
- Paraskevopoulos, E., Kuchenbuch, A., Herholz, S. C., & Pantev, C. (2012b). Musical expertise induces audiovisual integration of abstract congruency rules. *Journal of Neuroscience*, *32*, 18196–18203.
- Pardo, J., & Fox, P. (1991). Localization of a human system for sustained attention by positron emission tomography. *Nature*, *349*, 61–64.
- Pascual-Marqui, R., & Michel, C. (1994). Low resolution electromagnetic tomography: A new method for localizing electrical activity in the brain. *International Journal of Psychophysiology*, *18*, 49–65.
- Passingham, R. E., & Toni, I. (2001). Contrasting the dorsal and ventral visual systems: Guidance of movement versus decision making. *NeuroImage*, *14*, S125–S131.
- Plank, T., Rosengarth, K., Song, W., Ellermeier, W., & Greenlee, M. W. (2011). Neural correlates of audiovisual object recognition: Effects of implicit spatial congruency. *Human Brain Mapping*, *33*, 797–811.
- Ramnani, N., & Owen, A. M. (2004). Anterior prefrontal cortex: Insights into function from anatomy and neuroimaging. *Nature Reviews Neuroscience*, *5*, 184–194.
- Rusconi, E., Kwan, B., Giordano, B. L., Umiltà, C., & Butterworth, B. (2006). Spatial representation of pitch height: The SMARC effect. *Cognition*, *99*, 113–129.
- Rüsseler, J., Altenmüller, E., Nager, W., Kohlmetz, C., & Münte, T. F. (2001). Event-related brain potentials to sound omissions differ in musicians and nonmusicians. *Neuroscience Letters*, *308*, 33–36.
- Scholz, J., Klein, M. C., Behrens, T. E. J., & Johansen-Berg, H. (2009). Training induces changes in white-matter architecture. *Nature Neuroscience*, *12*, 1370–1371.
- Song, X.-W., Dong, Z.-Y., Long, X.-Y., Li, S.-F., Zuo, X.-N., Zhu, C.-Z., et al. (2011). REST: A toolkit for resting-state functional magnetic resonance imaging data processing. *PLoS One*, *6*, e25031.
- Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception & Psychophysics*, *73*, 971–995.
- Stewart, L. (2005). A neurocognitive approach to music reading. *Annals of the New York Academy of Sciences*, *1060*, 377–386.
- Stewart, L., Henson, R., Kampe, K., Walsh, V., Turner, R., & Frith, U. (2003). Brain changes after learning to read and play music. *NeuroImage*, *20*, 71–83.
- Tse, C.-Y., Tien, K.-R., & Penney, T. B. (2006). Event-related optical imaging reveals the temporal dynamics of right temporal and frontal cortex activation in pre-attentive change detection. *NeuroImage*, *29*, 314–320.
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, *15*, 273–289.
- Ungerleider, L., Courtney, S., & Haxby, J. (1998). A neural system for human visual working memory. *Proceedings of the National Academy of Sciences, U.S.A.*, *95*, 883–890.
- Urakawa, T., Inui, K., Yamashiro, K., & Kakigi, R. (2010). Cortical dynamics of the visual change detection process. *Psychophysiology*, *47*, 905–912.
- Waberski, T. D., Kreitschmann-Andermahr, I., Kawohl, W., Darvas, F., Ryang, Y., Gobbel, R., et al. (2001).

- Spatio-temporal source imaging reveals subcomponents of the human auditory mismatch negativity in the cingulum and right inferior temporal gyrus. *Neuroscience Letters*, 308, 107–110.
- Walker, P., Bremner, J., Mason, U., Spring, J., Mattok, K., Slater, A., et al. (2010). Preverbal infants' sensitivity to synaesthetic cross-modality correspondences. *Psychological Science*, 21, 21–25.
- Werner, S., & Noppeney, U. (2010). Superadditive responses in superior temporal sulcus predict audiovisual benefits in object categorization. *Cerebral Cortex*, 20, 1829.
- Yumoto, M., Uno, A., Itoh, K., Karino, S., Saitoh, O., Kaneko, Y., et al. (2005). Audiovisual phonological mismatch produces early negativity in auditory cortex. *NeuroReport*, 16, 803–806.
- Zatorre, R. J., & Gandour, J. T. (2008). Neural specializations for speech and pitch: Moving beyond the dichotomies. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences*, 363, 1087–1104.
- Zimmerman, E., & Lahav, A. (2012). The multisensory brain and its ability to learn music. *Annals of the New York Academy of Sciences*, 1252, 179–184.