Behavioral/Systems/Cognitive

# Musical Expertise Induces Audiovisual Integration of Abstract Congruency Rules

**Evangelos Paraskevopoulos,**[1] **Anja Kuchenbuch,**[1] **Sibylle C. Herholz,**[2] **and Christo Pantev**[1]

[1]Institute for Biomagnetism and Biosignalanalysis, University of Münster, D-48149 Münster, Germany, and [2]Montreal Neurological Institute, McGill University, Montreal, Quebec H3A 2B4, Canada

Perception of everyday life events relies mostly on multisensory integration. Hence, studying the neural correlates of the integration of multiple senses constitutes an important tool in understanding perception within an ecologically valid framework. The present study used magnetoencephalography in human subjects to identify the neural correlates of an audiovisual incongruency response, which is not generated due to incongruency of the unisensory physical characteristics of the stimulation but from the violation of an abstract congruency rule. The chosen rule—"the higher the pitch of the tone, the higher the position of the circle"—was comparable to musical reading. In parallel, plasticity effects due to long-term musical training on this response were investigated by comparing musicians to non-musicians. The applied paradigm was based on an appropriate modification of the multifeatured oddball paradigm incorporating, within one run, deviants based on a multisensory audiovisual incongruent condition and two unisensory mismatch conditions: an auditory and a visual one. Results indicated the presence of an audiovisual incongruency response, generated mainly in frontal regions, an auditory mismatch negativity, and a visual mismatch response. Moreover, results revealed that long-term musical training generates plastic changes in frontal, temporal, and occipital areas that affect this multisensory incongruency response as well as the unisensory auditory and visual mismatch responses.

## Introduction

Perception of everyday life events, such as watching and listening to a movie, relies mostly on multisensory integration. Hence, studying the neural correlates of the integration of multiple senses constitutes an important tool in understanding perception within an ecologically valid framework. Music notation reading combines visual, auditory, and motoric information (Zatorre, 2005) and therefore is a very useful process for studying multisensory integration (Schön et al., 2002). Musicians' long-term training modifies relevant cortical functions and representations (Jäncke, 2009) and thus their training in music reading provides a solid basis for studying multisensory plasticity.

Indications for genuine multisensory phenomena come from several different fields and neurophysiological studies have made important contributions to the study of multisensory integration (Besle et al., 2009). A common approach for studying multisensory integration is the manipulation of congruency between unisensory information and multisensory stimuli. This method has the advantage that attention is balanced across the different sensory conditions and thus attention-related issues that may affect the interpretation of cross-modal interaction are overcome (Calvert and Thesen, 2004).

Mismatch negativity (MMN) is an event-related component known to assess discrimination of novel sound events from an expected input. Typically, it is generated in auditory cortex and in response to a deviant sound within a stream of standard sounds. MMN latency ranges between 100 and 250 ms after the onset of the deviant stimulus (Näätänen, 1995). MMN has been shown to also be present in the visual modality (Pazo-Alvarez et al., 2003). In an audiovisual framework, the presence of an MMN response to incongruent audiovisual stimuli indicates violation of the expectancy that the stimuli of the unisensory modalities are congruent, and thus can be interpreted as objective evidence that the information of the visual and auditory stimuli have been integrated (Stekelenburg et al., 2004; Yumoto et al., 2005).

Recently, Lee and Noppeney (2011) investigated the temporal window of audiovisual integration in a study comparing musicians and non-musicians. Their results indicated that practicing piano shapes automatic audiovisual temporal binding by a context-specific neural mechanism selectively for music, but not for speech.

The goal of the present study was to investigate the influence of long-term musical training on multisensory processing. We used magnetoencephalography (MEG) to study the integration of auditory and visual stimuli when the congruency (or incongruency) of the stimuli does not rely on physical characteristics of the stimulation, such as temporal synchronization (Lee and Noppeney, 2011), but on an abstract convention such as the rule behind musical reading: "the higher the pitch of the tone, the higher the position of the circle". For this purpose, we used a multifeature oddball paradigm with audiovisual, auditory, and visual deviances combined within the context of a simplified mu-
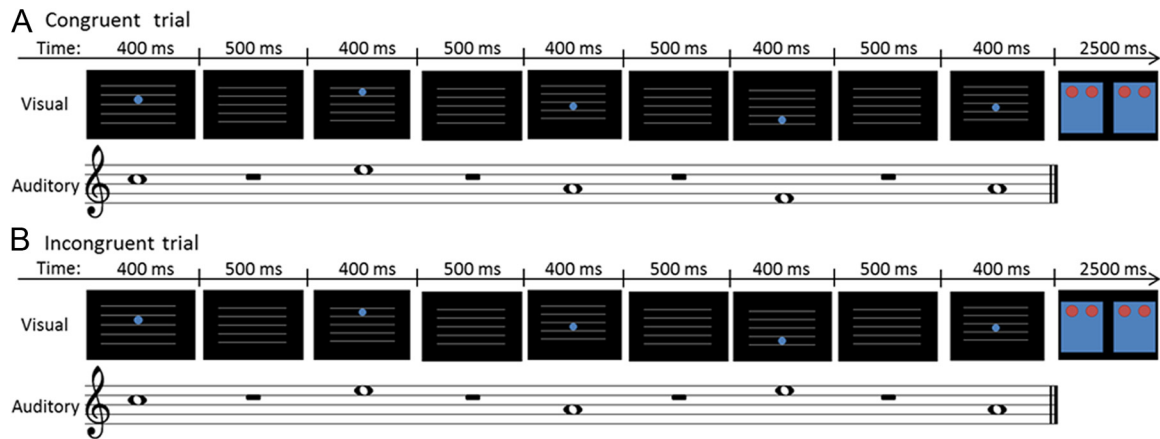
**Figure 1.** Examples of audiovisual congruent and incongruent trials. **A**, Congruent trial (that also served as auditory and visual standard). **B**, Incongruent trial. The line "Time" represents the duration of the presentation of the auditory and visual part of the stimulus. The last picture of each trial represents the intertrial stimulus in which subjects had to answer if the trail was congruent or incongruent.

sic reading task. Our hypothesis was that musicians would show stronger activation in an audiovisual incongruency response and that this response would be differentiated from unisensory MMNs, revealing an enhanced responsiveness in a genuine audiovisual process, which would indicate plasticity effects of musical training on multisensory processing.

## Materials and Methods

*Subjects.* The sample of the present study consisted of 30 subjects: 15 musicians and 15 non-musicians. Musicians (mean age = 25.66 years; SD = 2.76 years; 7 males) were students of the Music Conservatory in Münster (mean musical training = 18.06 years; SD = 4.76 years) studying various instruments (7 piano, 3 violin, 2 guitar, 1 accordion, 1 flute, and 1 cello). Non-musicians (mean age = 26.46 years; SD = 2.82 years; 9 males) had not received any formal musical education apart from compulsory school lessons. All subjects had normal hearing as evaluated by clinical audiometry and were right-handed. Evaluation of handedness was based on a score >40 in the Edinburgh Handedness Inventory (Oldfield, 1971). Subjects 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.* Four different categories of stimuli (audiovisual congruent, audiovisual incongruent, auditory deviant, and visual deviant) 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 (Fig. 1) similar to the staff lines in music notation. A blue circle was then placed in the middle of the horizontal direction and in one of the four spaces between the lines. The five tone melodies 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 ms and 10 ms rise and decay time (44,100 kHz, 16 bit). The interstimulus interval between each tone was 500 ms and the total duration of each melody was 4 s. Eight different melodies were prepared for each video category and the first tone of all melodies was C5.

For the audiovisual congruent stimuli the melodies were combined with the corresponding image according to the rule "the higher the pitch, the higher the position of the disk," following, thus, the music notation system. For each tone played, the blue circle appeared at the appropriate location at the same time and for the same duration as the tone. The white lines were continuously present.

The audiovisual incongruent stimuli were prepared along the same principle as the congruent ones, except that one of the tones of the melodies was paired with an image that did not correspond with the tone

according to the above-mentioned rule, thus producing an audiovisual incongruency. The difference always violated the contour of the melody and not simply the interval. The incongruency was never in the first tone–image pair but was equally likely to appear in each of the other four places (Fig. 1). The auditory (timbre) deviant stimuli were again prepared along the same principles as the congruent audiovisual ones, 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. Thus, the audiovisual correspondence of these stimuli was congruent (since the pitch height was correct) but the auditory input of this tone was deviant to all other tones due to the different timbre. The deviance was never in the first tone–image pair but was equally likely to appear in each of the other four places. For the visual (color) deviant stimuli, the same procedure was used as in the audiovisual congruent ones, but the blue circle of one of the images was replaced with a red circle, creating a condition that was congruent in terms of audiovisual correspondence but deviant in terms of the visual input. Again, the deviance was never in the first tone–image pair but was equally likely to appear in each of other four places.

In total, there were four conditions of stimuli, all providing audiovisual stimulation: Condition 1 included no mismatches and is therefore referred to as standard, Condition 2 included an audiovisual incongruency (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).

*MEG recordings.* Evoked magnetic fields were recorded with a 275-channel whole-head system (OMEGA; CTF Systems) in a magnetically shielded room. Data were acquired continuously during each presentation block with a sampling rate of 600 Hz. Subjects were seated upright, and their head position was comfortably stabilized with pads inside the dewar. The auditory part of the stimuli was delivered via 60-cm-long silicon tubes at 60 dB SL above the individual hearing threshold, which 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 ~150 cm away from the subject's nasion. The monitor was run at 60 Hz and a spatial resolution of 1280 × 1024 pixels. The viewing angle of the stimuli ranged from −3.86° to 3.86° in the horizontal direction and from −1.15° to 1.15° in the vertical direction. The recording was synchronized to: (1) the presentation of all tones of the audiovisual congruent trials that served as standards for all modalities, (2) the incongruent tone of the audiovisual incongruent trials serving as deviants in the audiovisual modality, (3) the deviant timbre tone of the auditory deviant trials serving as deviants in the auditory modality, and (4) the deviant color circle of the visual deviant trials serving as deviants for the visual modality. This resulted in an

incongruent (deviant) to congruent (standard) ratio of 20% for all modalities.

*Design.* Videos from all four conditions were randomly presented in one block. This block consisted of 32 videos of each video category randomly interleaved, creating a multifeatured oddball paradigm (Näätänen et al., 2004) appropriately adapted for a multisensory experiment. Within 2.5 s after each video was presented, subjects had to answer via button press if the video was congruent or incongruent (right hand) and if there was a tone sounding differently than all others or a circle of a different color (left hand). Thus, there were two buttons per hand. During this intertrial interval, an image was presented to the subjects to remind them which button represented each answer. Instructions for the task along with one video example of each category were given to the participants prior the beginning of the MEG recordings. Subjects were exposed to four blocks of stimuli, lasting ~14.5 min each, with short breaks in between. The total amount of stimuli for each category was 128.

*Data analysis.* The Brain Electrical Source Analysis software (BESA Research, version 5.3.7; Megis Software) was used to process the MEG data. The recorded data were separated in epochs of 700 ms, including a prestimulus interval of 200 ms. 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 ms. Averages of all four runs were computed separately for the congruent and the incongruent stimuli of the audiovisual modality and for the deviants of the auditory and visual modalities.

Current density reconstructions (CDR) were calculated on the neural responses of each subject for each stimulus category (congruent audiovisual, incongruent audiovisual, auditory deviant, visual deviant) using the low-resolution brain electromagnetic tomography (LORETA) method (Pascual-Marqui et al., 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 to the surface of the cortex. This method has been used successfully for the mapping of MMN (Waberski et al., 2001; Marco-Pallarés et al., 2005) 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 (130–170 ms). The appropriate time window was determined so as to be within the range of 110–250 ms, which is the typical latency window of a MMN, and to include the rising slope and the peak of the grand average global field power of the responses within this time range. This procedure is common in the MMN literature (Näätänen et al., 1989) and the specific time window is often chosen (Rüsseler et al., 2001; Tse et al., 2006). 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 (MNI) template. The images were smoothed and their intensities normalized by convolving an isotropic Gaussian kernel with 7 mm full width half-maximum through Besa's smoothing utility.

Statistical Parametric Mapping 8 (SPM8, http://www.fil.ion.ucl.ac.uk/spm) was used for the statistical analysis of the CDRs. Specifically, using the second level of analysis of SPM, a separate flexible factorial model was designed for each modality (audiovisual, auditory, and visual) to explore the main effect of condition (mismatch response) and the group × condition interaction. The flexible factorial model is SPM's equivalent analysis to a 2 × 2 mixed-model ANOVA with between-subject factor group and within-subject factor condition. The factors included in the analysis were subject (to control for the repeated measures), group (musicians and non-musicians), and condition (standard and deviant). Results were then 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 1% level 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, $p <$ 0.005 uncorrected) along with the cluster size (threshold size, >172 voxels), thereby controlling for multiple comparisons.

Additionally, to present the difference of congruent and incongruent conditions in the specific time window, unrestricted from the assump-

tions of the specific source analysis model, we conducted a statistical analysis of the fluctuation of the magnetic field in sensor space using BESA statistics. A nonparametric permutation test for paired samples was applied (1000 permutations) using a cluster alpha level of $p <$ 0.05; the distance indicating neighbor sensors was set at 4 cm. In this analysis, the data of the two groups were combined to reach a high level of signal-to-noise ratio.

# Results

## Behavioral responses

The discriminability index, $d'$, was calculated for the analysis of the behavioral responses. Both musicians and non-musicians scored significantly higher than the chance level [musicians: audiovisual condition: 2.83 ($t_{(14)} = 33.28$; $p < 0.000$); auditory condition: 3.71 ($t_{(14)} = 12.17$; $p < 0.000$); visual condition ($t_{(14)} = 16.57$; $p < 0.000$); non-musicians: audiovisual condition: 1.95 ($t_{(14)} = 8.5$; $p < 0.000$); auditory condition: 2.97 ($t_{(14)} = 10.7$; $p < 0.000$); visual condition: 3.95 ($t_{(14)} = 16.82$; $p < 0.000$)]. Moreover a 2 × 3 mixed-model ANOVA with between-subject factor group (musicians and non-musicians) and within-subject factor modality (audiovisual, auditory, and visual) revealed a main effect of modality ($F_{(2,27)} = 56.08$ $p < 0.000$), and an interaction of group × modality ($F_{(2,27)} = 3.56$ $p < 0.05$). Consequently, an independent samples $t$ test for each modality was calculated as *post hoc* comparison. After Bonferroni correction, the significance threshold for the *post hoc* tests was set to 0.017. The *post hoc* comparisons revealed that musicians scored significantly higher than non-musicians in the audiovisual modality ($t_{(28)} = 3.56$; $p < 0.001$) while no significant difference was observed in the auditory and visual modalities.

## MEG results

### Audiovisual modality

*Audiovisual incongruency response generators.* The main effect of condition of the statistical analysis for the audiovisual incongruency response revealed a complex network of generators. Specifically, effects were found in the anterior cingulate cortex (ACC; peak coordinates: $x = -10$, $y = 38$, $z = 10$; $t_{(28)} = 4.57$; cluster size = 3927 voxels; $p < 0.001$ AlphaSim corrected) extending to the superior frontal gyrus; a right frontal region (peak coordinates: $x = 44$, $y = 48$, $z = 18$; $t_{(28)} = 3.38$; cluster size = 268 voxels; $p < 0.001$ AlphaSim corrected) in the left inferior frontal gyrus (IFG; peak coordinates: $x = -52$, $y = 18$, $z = 12$; $t_{(28)} = 3.20$; cluster size = 210 voxels; $p < 0.001$ AlphaSim corrected), one in the right cingulate gyrus (peak coordinates: $x = 0$, $y = -2$, $z = 48$; $t_{(28)} = 3.25$; cluster size = 326 voxels; $p < 0.001$ AlphaSim corrected), and one in the right supramarginal gyrus (peak coordinates: $x = 60$, $y = -32$, $z = 40$; $t_{(28)} = 3.17$; cluster size = 210 voxels; $p < 0.001$ AlphaSim corrected). The above-mentioned results are summarized in Table 1 and the statistical map is presented in Figure 3. All anatomical regions are defined in Talairach space using TalairachClient (http://www.talairach.org/) after the transformation of SPM's MNI coordinates in Talairach space using icbm2tal (http://brainmap.org/icbm2tal/).

The statistical analysis of the fluctuation of the magnetic field in sensor space revealed that the two conditions differed significantly in four clusters of activity: two positive and two negative. Specifically, a negative cluster extending from right frontal to right temporal and right occipital sensors ($p = 0.000$), a positive cluster in the left frontal and temporal sensors ($p = 0.000$), a negative cluster in left central parietal sensors ($p = 0.001$), and a positive cluster in right central parietal sensors ($p = 0.001$). The results of this analysis are presented in Figure 2.

**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 $t_{(28)}$ value | Cluster size |
| | | $x$ | $y$ | $z$ | | |
|---|---|---|---|---|---|---|
| Audiovisual | Anterior cingulate cortex | −10 | 38 | 10 | 4.57 | 3927 |
| | Right middle frontal gyrus | 44 | 48 | 18 | 3.38 | 268 |
| | Right cingulate gyrus | 0 | −2 | 48 | 3.25 | 326 |
| | Left inferior frontal gyrus | −52 | 18 | 12 | 3.20 | 210 |
| | Right supramarginal gyrus | 60 | −32 | 40 | 3.17 | 210 |
| Auditory | Right superior frontal gyrus | 38 | 48 | 32 | 4.99 | 1947 |
| | Left middle frontal gyrus | −28 | 30 | 42 | 4.18 | 7527 |
| | Right superior temporal gyrus | 52 | −8 | 10 | 3.43 | 4561 |
| Visual | Left fusiform gyrus | −38 | −76 | −18 | 3.73 | 1257 |
| | Right fusiform gyrus | 50 | −60 | 16 | 3.55 | 452 |
| | Cingulate cortex | 6 | −4 | 46 | 3.15 | 547 |
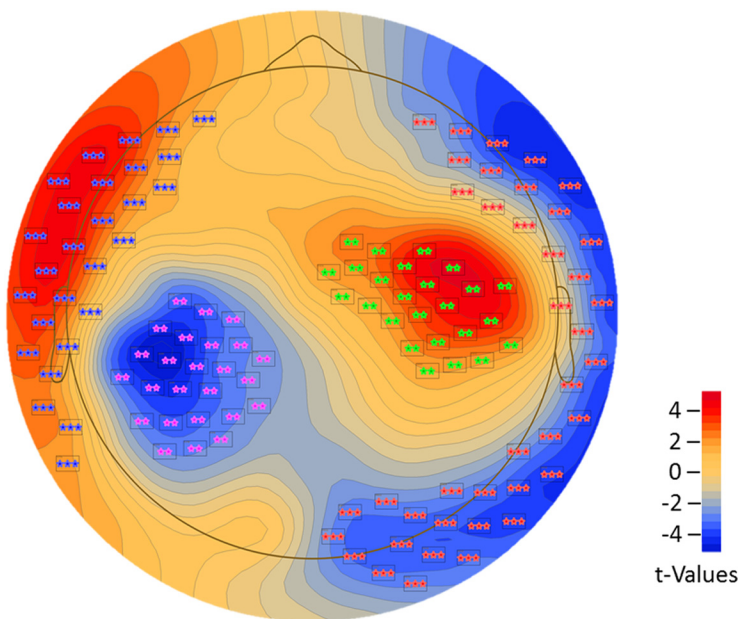


**Figure 2.** Distribution of the magnetic field of the audiovisual incongruency response in sensor space. The stars indicate clusters of sensors where the difference between the congruent and the incongruent condition is significant according to the permutation tests. The three red or three blue stars indicate significance of $p = 0.000$; while the two green or two violet stars indicate significance of $p = 0.001$.

*Musicians versus non-musicians comparison.* The statistical analysis of the audiovisual modality revealed a significant interaction of group × condition, thereby showing a differentiated response on the two groups. This interaction was investigated using a $t$-contrast to examine the specific direction of the effect. The result was located in three clusters of activity: one in the right superior frontal gyrus (SFG; peak coordinates: $x = 6, y = 20, z = 68$; $t_{(28)} = 4.55$; cluster size = 1143 voxels; $p < 0.001$ AlphaSim corrected); one in the right superior temporal gyrus (STG; peak coordinates: $x = 60, y = -6, z = 12$; $t_{(28)} = 3.43$; cluster size = 861 voxels; $p < 0.001$ AlphaSim corrected); and one in the visual cortex, right lingual gyrus (LG; peak coordinates: $x = 4, y = -94, z = -16$; $t_{(28)} = 3.60$; cluster size = 204 voxels; $p < 0.001$ AlphaSim corrected). Thereby, it was revealed that musicians showed significantly greater activity in these regions when confronted with an audiovisual incongruency. The opposite contrast of non-musicians showing greater activity than musicians did not reveal a significant activation. The statistical map of the comparison of musicians to non-musicians is presented in Figure 3.

*Auditory modality*
*MMN generators.* The main effect of condition of the statistical analysis for the auditory MMN response revealed three clusters of activity. Specifically, one cluster of activity was located in the right temporal cortex (peak coordinates: $x = 52, y = -8, z = -10$; $t_{(28)} = 3.43$; cluster size = 4561 voxels; $p < 0.001$ AlphaSim corrected) and two clusters were located bilaterally in frontal regions (peak coordinates: right hemisphere: $x = 38, y = 48, z = 32$; $t_{(28)} = 4.99$; cluster size = 1947 voxels; $p < 0.001$ AlphaSim corrected; left hemisphere: $x = -28, y = 30, z = 42$; $t_{(28)} = 4.18$; cluster size = 7527 voxels; $p < 0.001$ AlphaSim corrected). The above-mentioned results are summarized in Table 1 and the statistical map is presented in Figure 3.

*Musicians versus non-musicians comparison.* The statistical analysis of the auditory MMN revealed a significant interaction of group × condition, thereby showing increased activation in the group of musicians. This interaction was located in two clusters of activity: one in the anterior part of the STG (peak coordinates: $x = 26, y = 10, z = -36$; $t_{(28)} = 3.65$; cluster size = 671 voxels; $p < 0.001$ AlphaSim corrected) and one in the ACC (peak coordinates: $x = -14, y = 64, z = 28$; $t_{(28)} = 3.06$; cluster size = 742 voxels; $p < 0.001$ AlphaSim corrected). The opposite contrast of non-musicians showing greater activity than musicians did not reveal a significant activation. The statistical map of the comparison of musicians to non-musicians is presented in Figure 4.

*Visual modality*
*Mismatch response generators.* The main effect of condition of the statistical analysis for the visual mismatch response revealed three clusters of activity. Specifically, two of the clusters were located bilaterally in the left and right fusiform gyrus (FG; peak coordinates: left hemisphere: $x = -38, y = -76, z = -18$; $t_{(28)} = 4.39$; cluster size = 1257 voxels; $p < 0.001$ AlphaSim corrected; right hemisphere: $x = 50, y = -60, z = -16$; $t_{(28)} = 3.55$; cluster size = 452 voxels; $p < 0.001$ AlphaSim corrected) and one in the cingulate cortex (peak coordinates: $x = 6, y = -4, z = 46$; $t_{(28)} = 3.15$; cluster size = 547 voxels; $p < 0.001$ AlphaSim corrected). The
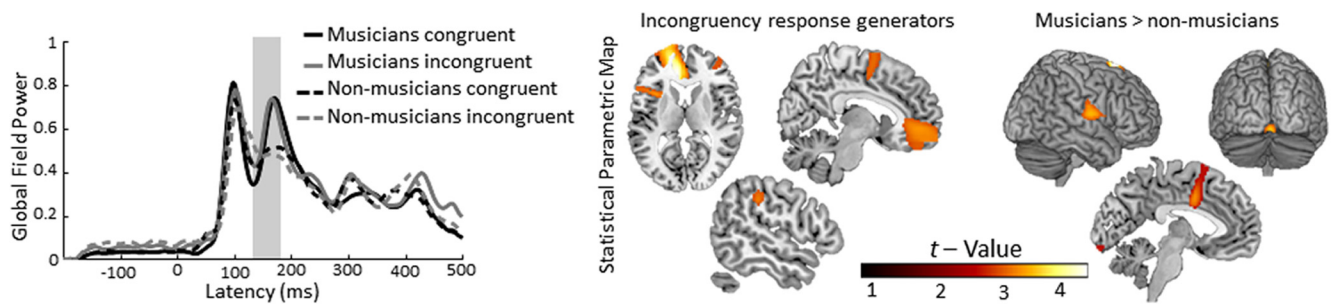
## Audio-visual Modality



**Figure 3.** Right, Statistical parametric maps of the audiovisual incongruency response and the musicians to non-musicians comparison, as revealed by the flexible factorial model. Threshold: AlphaSim corrected at $p < 0.001$ by taking into account peak voxel significance (threshold $p < 0.001$ uncorrected) and cluster size (threshold size, $>172$ voxels). Left, Grand average global field power for congruent (black lines) and incongruent (gray lines) response for musicians (continuous lines) and non-musicians (dashed lines). Gray bar, Time interval where the analysis was performed.
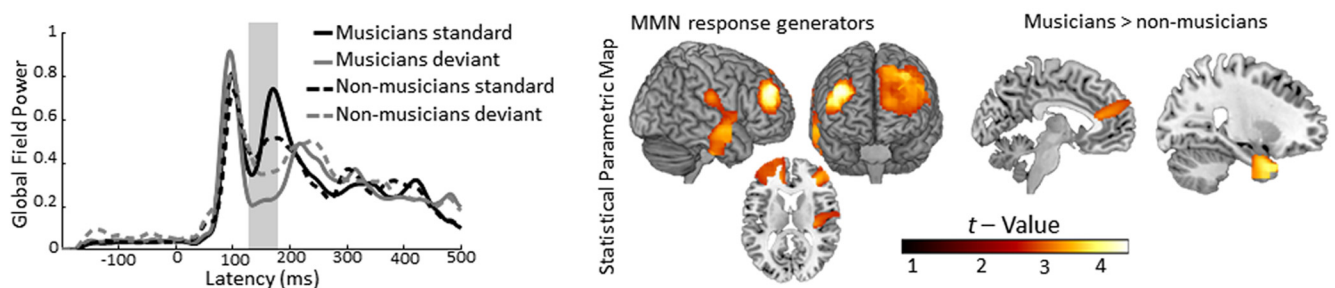
## Auditory modality



**Figure 4.** Right, Statistical parametric maps of the auditory MMN response and the musicians to non-musicians comparison as revealed by the flexible factorial model. Threshold: AlphaSim corrected at $p < 0.001$ by taking into account peak voxel significance (threshold $p < 0.001$ uncorrected) and cluster size (threshold size, $>172$ voxels). Left, Grand average global field power for standard (black lines) and deviant (gray lines) response for musicians (continuous lines) and non-musicians (dashed lines). Gray bar, Time interval where the analysis was performed.
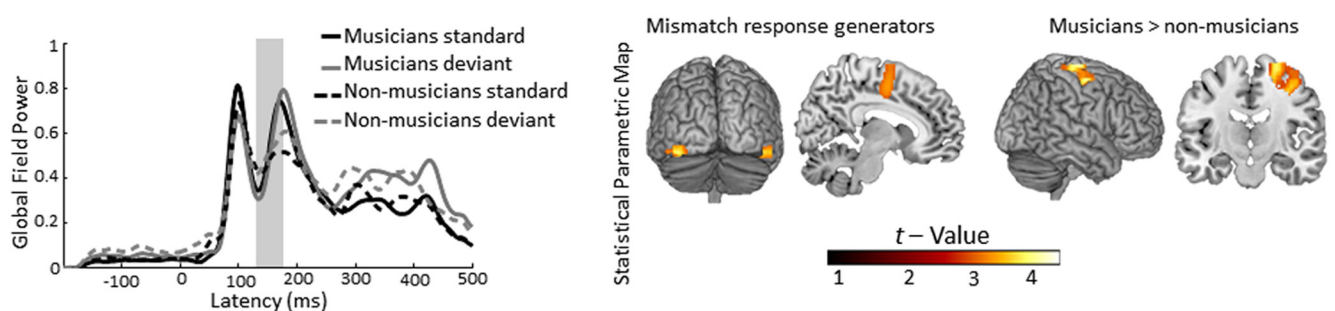
## Visual modality



**Figure 5.** Right, Statistical parametric maps of the visual mismatch response and the musicians to non-musicians comparison as revealed by the flexible factorial model. Threshold: AlphaSim corrected at $p < 0.001$ by taking into account peak voxel significance (threshold $p < 0.001$ uncorrected) and cluster size (threshold size, $>172$ voxels). Left, Grand average global field power for standard (black lines) and deviant (gray lines) response for musicians (continuous lines) and non-musicians (dashed lines). Gray bar, Time interval where the analysis was performed.

above-mentioned results are summarized in Table 1 and the statistical map is presented in Figure 5.

*Musicians versus non-musicians comparison.* The statistical analysis of the visual mismatch response revealed a significant interaction of group × condition, thereby showing increased activation in the group of musicians. This interaction was located in the right middle frontal gyrus (MFG; peak coordinates: $x = 24$, $y = -18$, $z = 68$; $t_{(28)} = 3.85$; cluster size = 982 voxels; $p < 0.001$ AlphaSim corrected). The opposite contrast of non-musicians

showing greater activity than musicians did not reveal a significant activation. The statistical map of the comparison of musicians to non-musicians is presented in Figure 5 and Table 2.

## Discussion

The present study used magnetoencephalography to identify the neural correlates of an audiovisual incongruency response generated due to violations of an abstract congruency rule. The cho-

**Table 2. Location of activity in musicians versus non-musicians comparison**

| Modality | Location of activation | Coordinates | | | Peak voxel $t_{(28)}$ value | Cluster size |
| --- | --- | --- | --- | --- | --- | --- |
| | | x | y | z | | |
| Audiovisual | Right superior frontal gyrus | 0 | 10 | 39 | 4.55 | 1143 |
| | Right superior temporal gyrus | 58 | −9 | 8 | 3.43 | 861 |
| | Right lingual gyrus | 2 | −87 | −14 | 3.60 | 204 |
| Auditory | Right superior temporal gyrus | 26 | 6 | −37 | 3.27 | 671 |
| | Left anterior cingulate gyrus | −4 | 42 | 17 | 3.13 | 742 |
| Visual | Right middle frontal gyrus | 24 | −18 | 68 | 3.85 | 982 |

sen rule was comparable to musical reading: "the higher the pitch of the tone—the higher the position of the circle." Plasticity effects of long-term musical training on this response were investigated by comparing musicians to non-musicians. The applied paradigm was based on a modification of the multifeatured oddball paradigm (Näätänen et al., 2004) combining an audiovisual incongruent condition, an auditory MMN condition, and a visual mismatch condition within one run. Results indicated the presence of an audiovisual incongruency response generated mainly in frontal regions. Moreover, musicians showed increased activity in frontal, temporal, and occipital areas as a response to multisensory incongruencies as well as unisensory auditory and visual mismatch responses.

The behavioral results imply that the task was not a solely musical one. Instead, it was an appropriate task to investigate audiovisual integration in both musicians and non-musicians. The fact that non-musicians scored well above chance level indicates that the rule was well understood and the incongruency was evident for both groups. Indeed, behavioral results supporting cross-modal congruency effects between the height of a pitch and the height of a visual stimulus have also been demonstrated using classification tasks in psychophysical studies testing non-musicians (Evans and Treisman, 2010; Spence, 2011). Despite the above-mentioned result, musicians scored significantly better in audiovisual modality, revealing the predicted effect of their musical training.

MEG results of the audiovisual modality clearly revealed an incongruency response. A complex cortical network contributed to this response. The discrimination of congruent from incongruent trials relies on audiovisual integration but also on working memory, since one has to keep in mind all five tone–circle pairs presented and maintain attention on the comparison of the audiovisual stimuli to the abstract rule. The ACC has been shown previously to be directly connected with audiovisual integration (Benoit et al., 2010) as well as attention load (Pardo et al., 1991; Bush et al., 2000). The left IFG is a structure with activity correlated with many different processes, but nevertheless it is also well correlated with working memory (Courtney et al., 1997) and has been linked with music reading in a previous fMRI study (Stewart et al., 2003). Similar to our current results, previous studies using congruent and incongruent audiovisual stimuli (McGurk effect) have also identified the right supramarginal gyrus as a structure contributing to the identification of incongruent stimuli (Jones and Callan, 2003; Benoit et al., 2010). Activation in supramarginal gyrus was also correlated with the translation of three different notation systems to music in an fMRI study by Schön et al. (2002) and with music reading in an fMRI study by Stewart et al. (2003), although in this study it was the left supramarginal gyrus that revealed activation. Convincingly, supramarginal gyrus is a structure highly correlated with audiovisual integration and especially with music reading, although its precise role remains unclear. The fact that we used an abstract discrimination rule might be the reason that activation was observed in more frontal regions, typically involved in higher cognitive functions (Badre and Wagner, 2004) in contrast to the study of Lee and Noppeney (2011) that investigated the temporal integration window of audiovisual stimuli. Thus, the integration of the audiovisual stimuli based on this abstract rule provides new insights to relevant audiovisual correspondences revealed in previous psychophysical (Parise and Spence, 2009) and fMRI (Sadaghiani et al., 2009) studies.

Musical training relies very strongly on audio–visual integration, particularly when reading music notation. The comparison of musicians to non-musicians in the audiovisual response revealed plasticity effects that can be attributed to long-term musical training. Specifically, increased activation in the group of musicians was found in the right secondary auditory cortex (STG), the visual cortex (LG), and the superior frontal gyrus. The auditory and the visual activations are highly correlated with the experimental task and SFG may be interpreted as a contribution of motor association areas originating from the audio–visual–motoric binding of musical notation that musicians have highly trained by practicing their instrument (Stewart et al., 2004; Schön and Besson, 2005).

Results of the auditory modality indicate a MMN response. The global field power of the standard condition (audiovisual congruent) reveals larger overall activation than the auditory deviant condition. This can be attributed to the fact that in the standard condition, subjects have to integrate the auditory and the visual information, while in the auditory deviant condition, subjects can rely solely on the auditory information. Indeed, multisensory integration is known to produce greater activity than unisensory processing (Stein and Meredith, 1993). Despite the larger activation in the global field power of the standard condition, regions typically associated with auditory MMN (STG) revealed increased activity during the auditory deviant presentation. The right lateralization of the response is typical for musical stimuli and consistent with the results of a PET study that compared musical and phonetic MMNs (Tervaniemi et al., 2000).

Furthermore, we would like to note the prominent frontal activation found in the auditory condition. Although a frontal component of auditory MMN has been described and attributed to attention changes related to the presentation of deviant stimuli (Näätänen et al., 1978; Alho et al., 1994), this was found using electroencephalography (EEG) and not MEG. Therefore, the frontal component of MMN was thought to be generated by radial sources, which are hard to detect with MEG. A study by Rinne et al. (2000) used both EEG and MEG to investigate auditory MMN and was able to show that this frontal activation was present only in the EEG data. Therefore, our interpretation of this frontal activation is that it does not reflect the typical frontal component of auditory MMN. Instead, consistently with the concept of attention changes due to deviant presentation, it may reflect atten-

tion switching between the multisensory condition (audiovisual) of our task and the unisensory one (auditory MMN).

The interaction of group × condition in the auditory MMN revealed that musicians showed increased activation in STG and ACC. As already noted, ACC is a structure linked to audio–visual integration (Benoit et al., 2010) as well as attention (Bush et al., 2000), and therefore it is expected that in a task where visual and auditory stimuli are simultaneously presented, this region is activated more strongly in musicians than in non-musicians. The activation of the anterior part of STG (temporal pole) was surprising. This region has been linked to passive music listening (Brown et al., 2004) and musical expertise (James et al., 2008), but it is also correlated with the coupling of perceptual characteristics of auditory stimulation to emotional processing (Olson et al., 2007) and especially with activations while listening to unpleasant music (Koelsch et al., 2006). In line with those findings, a tone of sawtooth waveform (deviant tones) sounds more harsh or unpleasant than a pure tone (standard tones) and a fast acoustical analysis of this stimulus may have evoked components related with the identification of this unpleasantness (Peretz, 2010). Recent evidence (James et al., 2008) suggest that musicians show increased responses to unpleasant musical stimuli compared with non-musicians, supporting our result.

Results of the visual modality revealed a mismatch response originating from bilateral FG and cingulate cortex. This activation pattern cannot be directly compared with recent MEG studies of visual mismatches (Urakawa et al., 2010; Paraskevopoulos et al., 2012) due to differences in the design and the complexity of the present paradigm. Nevertheless, the location of the activation seems to be well suited for a color mismatch. Specifically, FG is a region that has been directly connected with color perception (Hsu et al., 2011) as well as visual working memory (Ungerleider et al., 1998). The comparison of musicians versus non-musicians in the visual response revealed that musicians showed increased activation in MFG, a motor association area. This effect was also present in the musicians versus non-musicians comparison of the audiovisual modality response, indicating that it is related to the characteristics of the paradigm and not to a specific modality. Similar activations have been shown in other recent music reading studies (Stewart et al., 2004; Schön and Besson, 2005) and, therefore, this activation may be based on the long-term training of musicians in binding the musical notation with motoric actions when practicing their instrument.

In conclusion, the present study proposed a novel paradigm that combines unimodal and multimodal mismatches (and incongruencies). Results indicate an integrative audiovisual response based on the incongruency of an abstract rule that binds the auditory and visual inputs, and revealed its neural correlates. Moreover, results revealed that musicians showed an enhancement of this multisensory incongruency response, suggesting long-term training effects on neural correlates of multisensory integration.

# References

Alho K, Woods DL, Algazi A, Knight RT, Näätänen R (1994) Lesions of frontal cortex diminish the auditory mismatch negativity. Electroencephalogr Clin Neurophysiol 91:353–362. CrossRef Medline

Badre D, Wagner AD (2004) Selection, integration, and conflict monitoring: assessing the nature and generality of prefrontal cognitive control mechanisms. Neuron 41:473–487. CrossRef Medline

Benoit MM, Raij T, Lin FH, Jääskeläinen IP, Stufflebeam S (2010) Primary and multisensory cortical activity is correlated with audiovisual percepts. Hum Brain Mapp 31:526–538. CrossRef Medline

Besle J, Bertrand O, Giard MH (2009) Electrophysiological (EEG, sEEG, MEG) evidence for multiple audiovisual interactions in the human auditory cortex. Hear Res 258:143–151. CrossRef Medline

Brown S, Martinez Mj, Parsons LM (2004) Passive music listening spontaneously engages limbic and paralimbic systems. Neuroreport 15:2033–2037. CrossRef Medline

Bush G, Luu P, Posner Mi (2000) Cognitive and emotional influences in anterior cingulate cortex. Trends Cogn Sci 4:215–222. CrossRef Medline

Calvert GA, Thesen T (2004) Multisensory integration: methodological approaches and emerging principles in the human brain. J Physiol Paris 98:191–205. CrossRef Medline

Courtney SM, Ungerleider LG, Keil K, Haxby JV (1997) Transient and sustained activity in a distributed neural system for human working memory. Nature 386:608–611. CrossRef Medline

Evans KK, Treisman A (2010) Natural cross-modal mappings between visual and auditory features. J Vis 10:1–12. CrossRef Medline

Hsu NS, Kraemer DJ, Oliver RT, Schlichting ML, Thompson-Schill SL (2011) Color, context, and cognitive style: variations in color knowledge retrieval as a function of task and subject variables. J Cogn Neurosci 23:2544–2557. CrossRef Medline

James CE, Britz J, Vuilleumier P, Hauert CA, Michel CM (2008) Early neuronal responses in right limbic structures mediate harmony incongruity processing in musical experts. Neuroimage 42:1597–1608. CrossRef Medline

Jäncke L (2009) Music drives brain plasticity. F1000 Biol Rep 2009:1–78. CrossRef Medline

Jones JA, Callan DE (2003) Brain activity during audiovisual speech perception: an fMRI study of the McGurk effect. Neuroreport 14:1129–1133. CrossRef Medline

Koelsch S, Fritz T, V Cramon VD, Müller K, Friederici AD (2006) Investigating emotion with music: an fMRI study. Hum Brain Mapp 27:239–250. CrossRef Medline

Lee H, Noppeney U (2011) Long-term music training tunes how the brain temporally binds signals from multiple senses. Proc Natl Acad Sci U S A 108:E1441–E1450. CrossRef Medline

Marco-Pallarés J, Grau C, Ruffini G (2005) Combined ICA-LORETA analysis of mismatch negativity. Neuroimage 25:471–477. CrossRef Medline

Näätänen R (1995) The mismatch negativity: a powerful tool for cognitive neuroscience. Ear Hear 16:6–18. CrossRef Medline

Näätänen R, Gaillard AW, Mäntysalo S (1978) Early selective-attention effect on evoked potential reinterpreted. Acta Psychologica 42:313–329. doi:10.1016/0001-6918(78)90006-9. CrossRef

Näätänen R, Paavilainen P, Reinikainen K (1989) Do event-related potentials to infrequent decrements in duration of auditory stimuli demonstrate a memory trace in man? Neurosci Lett 107:347–352. CrossRef Medline

Näätänen R, Pakarinen S, Rinne T, Takegata R (2004) The mismatch negativity (MMN): towards the optimal paradigm. Clin Neurophysiol 115:140–144. CrossRef Medline

Oldfield RC (1971) The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9:97–113. CrossRef Medline

Olson IR, Plotzker A, Ezzyat Y (2007) The enigmatic temporal pole: a review of findings on social and emotional processing. Brain 130:1718–1731. CrossRef Medline

Paraskevopoulos E, Kuchenbuch A, Herholz SC, Pantev C (2012) Evidence for training-induced plasticity in multisensory brain structures: an MEG study. PLoS One 7:e36534. CrossRef Medline

Pardo JV, Fox PT, Raichle ME (1991) Localization of a human system for sustained attention by positron emission tomography. Nature 349:61–64. CrossRef Medline

Parise CV, Spence C (2009) "When birds of a feather flock together": synesthetic correspondences modulate audiovisual integration in nonsynesthetes. PloS One 4:e5664. CrossRef Medline

Pascual-Marqui RD, Michel CM, Lehmann D (1994) Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. Int J Psychophysiol 18:49–65. CrossRef Medline

Pazo-Alvarez P, Cadaveira F, Amenedo E (2003) MMN in the visual modality: a review. Biol Psychol 63:199–236. CrossRef Medline

Peretz I (2010) Towards a neurobiology of musical emotions. In: Handbook of music and emotion: theory, research, applications (Juslin P, Sloboda JA, eds.), pp 99–126. Oxford: Oxford UP.

Rinne T, Alho K, Ilmoniemi RJ, Virtanen J, Näätänen R (2000) Separate

time behaviors of the temporal and frontal mismatch negativity sources. Neuroimage 12:14–19. CrossRef Medline

Rüsseler J, Altenmüller E, Nager W, Kohlmetz C, Münte TF (2001) Event-related brain potentials to sound omissions differ in musicians and non-musicians. Neurosci Lett 308:33–36. CrossRef Medline

Sadaghiani S, Maier JX, Noppeney U (2009) Natural, metaphoric, and linguistic auditory direction signals have distinct influences on visual motion processing. J Neurosci 29:6490–6499. CrossRef Medline

Schön D, Besson M (2005) Visually induced auditory expectancy in music reading: a behavioral and electrophysiological study. J Cogn Neurosci 17:694–705. CrossRef Medline

Schön D, Anton JL, Roth M, Besson M (2002) An fMRI study of music sight-reading. Neuroreport 13:2285–2289. CrossRef Medline

Song XW, Dong ZY, Long XY, Li SF, Zuo XN, Zhu CZ, He Y, Yan CG, Zang YF (2011) REST: a toolkit for resting-state functional magnetic resonance imaging data processing. PLoS One 6:e25031. CrossRef Medline

Spence C (2011) Crossmodal correspondences: a tutorial review. Atten Percept Psychophys 73:971–995. CrossRef Medline

Stein BE, Meredith MA (1993) The merging of the senses. Cambridge: MIT.

Stekelenburg JJ, Vroomen J, de Gelder B (2004) Illusory sound shifts induced by the ventriloquist illusion evoke the mismatch negativity. Neurosci Lett 357:163–166. CrossRef Medline

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. CrossRef Medline

Stewart L, Walsh V, Frith U (2004) Reading music modifies spatial mapping in pianists. Percept Psychophys 66:183–195. CrossRef Medline

Tervaniemi M, Medvedev SV, Alho K, Pakhomov SV, Roudas MS, Van Zuijen TL, Näätänen R (2000) Lateralized automatic auditory processing of phonetic versus musical information: a PET study. Hum Brain Mapp 10:74–79. CrossRef Medline

Tse CY, Tien KR, Penney TB (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. CrossRef Medline

Ungerleider LG, Courtney SM, Haxby JV (1998) A neural system for human visual working memory. Proc Natl Acad Sci U S A 95:883–890. CrossRef Medline

Urakawa T, Inui K, Yamashiro K, Kakigi R (2010) Cortical dynamics of the visual change detection process. Psychophysiology 47:905–912. CrossRef Medline

Waberski TD, Kreitschmann-Andermahr I, Kawohl W, Darvas F, Ryang Y, Gobbelé R, Buchner H (2001) Spatio-temporal source imaging reveals subcomponents of the human auditory mismatch negativity in the cingulum and right inferior temporal gyrus. Neurosci Lett 308:107–110. CrossRef Medline

Yumoto M, Uno A, Itoh K, Karino S, Saitoh O, Kaneko Y, Yatomi Y, Kaga K (2005) Audiovisual phonological mismatch produces early negativity in auditory cortex. Neuroreport 16:803–806. CrossRef Medline

Zatorre R (2005) Music, the food of neuroscience? Nature 434:312–315. CrossRef Medline