Action expertise reduces brain activity for audiovisual matching actions: An fMRI study with expert drummers

Karin Petrini [a,*], Frank E. Pollick [a], Sofia Dahl [b], Phil McAleer [a], Lawrie McKay [c], Davide Rocchesso [d], Carl Haakon Waadeland [e], Scott Love [a], Federico Avanzini [f], Aina Puce [g]

[a] Department of Psychology, University of Glasgow, Glasgow, Scotland, UK
[b] Department of Media Technology, Aalborg University Copenhagen, Copenhagen, Denmark
[c] Netherlands Institute for Neuroscience (NIN), Amsterdam, The Netherlands
[d] Department of Art and Industrial Design, IUAV University of Venice, Venice, Italy
[e] Department of Music, Norwegian University of Science and Technology, Trondheim, Norway
[f] Department of Information Engineering, University of Padua, Padua, Italy
[g] Department of Psychological and Brain Sciences, Indiana University, Bloomington, USA

ARTICLE INFO

ABSTRACT

When we observe someone perform a familiar action, we can usually predict what kind of sound that action will produce. Musical actions are over-experienced by musicians and not by non-musicians, and thus offer a unique way to examine how action expertise affects brain processes when the *predictability* of the produced sound is manipulated. We used functional magnetic resonance imaging to scan 11 drummers and 11 age- and gender-matched novices who made judgments on point-light drumming movements presented with sound. In Experiment 1, sound was synchronized or desynchronized with drumming strikes, while in Experiment 2 sound was always synchronized, but the natural covariation between sound intensity and velocity of the drumming strike was maintained or eliminated. Prior to MRI scanning, each participant completed psychophysical testing to identify personal levels of synchronous and asynchronous timing to be used in the two fMRI activation tasks. In both experiments, the drummers' brain activation was *reduced* in motor and action representation brain regions when sound *matched* the observed movements, and was similar to that of novices when sound was mismatched. This reduction in neural activity occurred bilaterally in the cerebellum and left parahippocampal gyrus in Experiment 1, and in the right inferior parietal lobule, inferior temporal gyrus, middle frontal gyrus and precentral gyrus in Experiment 2. Our results indicate that brain functions in action-sound representation areas are modulated by multimodal action expertise.

© 2011 Elsevier Inc. All rights reserved.

Introduction

Music and dance have formed a significant part of human culture throughout history. Listening to music performed by a group of musicians, as well as dancing with others, relies on the ability to integrate sight with sound, as well as predicting what sound an action will produce and when. This is something that we can do with relatively little effort in a social setting. As non-expert observers, we also have the capacity to appreciate the performances of virtuoso musicians and dancers from their audiovisual output — when these individuals have spent years perfecting and honing their motor skills. Indeed, the majority of human actions are multisensory in nature, and our ability to predict the auditory consequence of others' actions relies on acquired knowledge gained from performing these actions ourselves.

That our brain produces similar neural activity for seen and performed actions is now well-known (e.g. Gallese et al., 1996; Rizzolatti et al., 1996; Decety and Grèzes, 1999; Nishitani and Hari, 2000; Buccino et al., 2001, 2004; Iacoboni, 2005; Haslinger et al., 2005; Calvo-Merino et al., 2005, 2006). These so-called 'mirror neurons' are not triggered only by visual stimuli, but also by auditory stimuli (Kohler et al., 2002; Keysers et al., 2003; Wilson et al., 2004). Based on these findings, the existence of a cross-modal neural system for action recognition has been suggested (Kohler et al., 2002; Keysers et al., 2003). A recent study by Lahav et al. (2007) examined the underlying mechanisms linking motor and auditory actions, and showed that frontoparietal regions were activated when listening to pieces of music listeners had both learned and played, but not when listening to other familiar music pieces. These data are consistent with findings in the visuomotor domain, where regions such as the dorsal premotor cortex, inferior parietal sulcus and cerebellum are activated when participants view actions that form part of their existing motor repertoire (Calvo-Merino et al., 2005, 2006).

Relatively few studies have investigated how visual and auditory information interact when recognizing the actions of others and how

* Corresponding author.
E-mail address: k.petrini@ucl.ac.uk (K. Petrini).

expertise with the represented actions influences the underlying brain processes (e.g. Musacchia et al., 2007; Haslinger et al., 2005; Hodges et al., 2005). Here, we asked the question of whether expertise for represented audiovisual actions will modulate brain activation in action-sound representation areas when the natural relation between sight and sound of familiar actions is disrupted. We thus aim at determining the basis of the higher perceptual and motor expertise, respectively, of drummers as compared to novices. To this end, we developed two different experiments, the first of which aimed to examine the role of perceptual and motor expertise when processing human actions for which the natural synchronization between movements and sounds was eliminated. In line with this, the second experiment examined the role of perceptual and motor expertise when processing human actions for which the natural covariation between movements' velocity and sounds' intensity was eliminated.

For a number of reasons, we chose to use the biological motion of drumming actions which were highly familiar to expert drummers and unfamiliar to non-musicians. Our first premise was that in music performance each visual action produces a sound (Zatorre et al., 2007). Drummers can better anticipate when the sound will occur based on the viewed action, compared with novices (Petrini et al., 2009b). Consequently, this makes drumming actions ideal stimuli for studying interactions between seeing and hearing when perceiving others' actions. Secondly, drumming actions are over-experienced by drummers, but not by non-musicians, allowing the study of brain activation differences which are driven by differences in action expertise (Calvo-Merino et al., 2005, 2006). Thirdly, music has been extensively used to study visuomotor (e.g. Stewart et al., 2003; Buccino et al., 2004; Vogt et al., 2007) and audiomotor (e.g. Parsons et al., 2005; Baumann et al., 2005, 2007; Bangert et al., 2006; Lahav et al., 2007) brain processes involved in action representation. Finally, point-light biological motion (Johansson, 1973) has been previously used to focus on brain processes driven by visual motion cues of actions (Saygin et al., 2004; Saygin, 2007), as well as by audiovisual integration (Brooks et al., 2007; Klin et al., 2009).

We therefore ran two event-related fMRI experiments in 11 drummers and 11 novices matched for age and gender. In Experiment 1, they evaluated animated displays where the correspondence between drumming action and resulting sound was eliminated by desynchronizing the two signals. In Experiment 2, the same participants were scanned while viewing animated displays where the temporal correspondence between the signals was maintained, but covariation between the drummer's movements and the resulting sound was eliminated (i.e. sound intensity did not covary with the velocity). We hypothesized that acquiring perceptual and motor expertise through practice with the portrayed audiovisual action would alter brain processes, and that this would result in different networks of brain areas being activated in Experiments 1 and 2, since the difference between matching and mismatching displays in the two experiments differentiates between sensory synchrony and sensory congruency.

Materials and methods

Participants

Eleven right-handed drummers (all males; age 35 ± 12 years) were studied. Seven of the 11 were presently either professional musicians and/or music teachers. The average number of years of drumming experience was 24 ± 11 years (range 13 to 45 years). All drummers were right-handed, as assessed by the Edinburgh Handedness Questionnaire (Oldfield, 1971).

We gender- and age-matched our novice participant group, relative to our expert group, to study 11 males (age 35 ± 11 years) with no previous drumming experience and, importantly, little or no

previous musical training. Eight of the novices were right-handed, as assessed by the Edinburgh Handedness Questionnaire (Oldfield, 1971).

Stimulus creation

Audiovisual stimuli were created in three distinct steps from initial motion capture data of a professional jazz drummer playing a swing groove beat (Petrini et al., 2009a,b), as briefly described below.

Step 1 consisted of converting the 3D movement coordinates of the drummer into point-light displays of drumming actions, using computer graphics. Step 2 consisted of converting the times and velocities of stick impact into a realistic stream of sound by using a naturalistic sound generation algorithm. Step 3 consisted of combining the visual point-light displays and the audio stream of drum beats by using video editing software.

Motion-capture data

Data were recorded from the movements of a professional jazz drummer playing a swing groove at 120 beats per minute, with the accent on the second beat (Waadeland, 2006). Markers were placed at six locations, including the tip of the drumstick, the level of the grip of the drumstick, and at the drummer's shoulder, elbow, wrist and hand.

The 3D coordinates of these marker locations were sampled at 240 Hz using a Proreflex 3D motion capture system. The sampled 3D motion capture data were first downsampled to 60 Hz. They were then converted into a series of 2D images using Matlab and Psychtoolbox routines (Brainard, 1997; Pelli, 1997) running under OSX and utilizing OpenGL graphics with anti-aliasing enabled. White disks (luminance: 85 cd/m²; diameter: 2 mm) on a black background (luminance: 0.12 cd/m²) represented the drummer's arm and drumstick (Fig. 1). The drum head was represented using a thick white line (Fig. 1), orientated 25° from horizontal (width: 2.2 cm; height: 2 mm; luminance: 85 cd/m²). The image sequences were saved as video using an AVI file format at a 60 Hz frame rate.

Auditory stimuli

The synthetic drumming sounds were obtained by an algorithm that took as input the times and velocities of a series of strikes, and outputted the simulated audio signal (Fontana et al., 2004). To obtain a very natural sound, the algorithm simulated the first 25 modes of a circular membrane. Both the time and impact velocity of a strike were derived by plotting the displacement and velocity of the drumstick tip marker against the time of the drummer performance, and selecting, for each impact, the frame at which the drumstick tip velocity changed from negative to positive (Dahl, 2004, pp. 765). To use only displacement and velocity perpendicular to the drumhead (Dahl, 2004, pp. 765), data were rotated to a coordinate frame, where horizontal was parallel to the drumhead and vertical perpendicular to the drumhead. These operations were performed on the 240 Hz, unfiltered displacements and velocities.

The resulting sounds were saved as audio files (WAV format) with a 15-second duration. The audio files were created to either correspond to the original movement data recordings, or to not correspond (see Fig. 1). The corresponding audio files were obtained by taking the impact times and impact velocities found in the original recorded movement and using these as input to the algorithm. For the non-corresponding audio files, the identical impact times were used, but the set of velocities was randomly scrambled. Thus, for the matching displays, there was a natural covariation between the original movements of the drummer and the resulting sound (e.g. fast strikes, loud sounds; slow strikes, soft sounds), while for the mismatching displays the timing of the strikes was identical, but
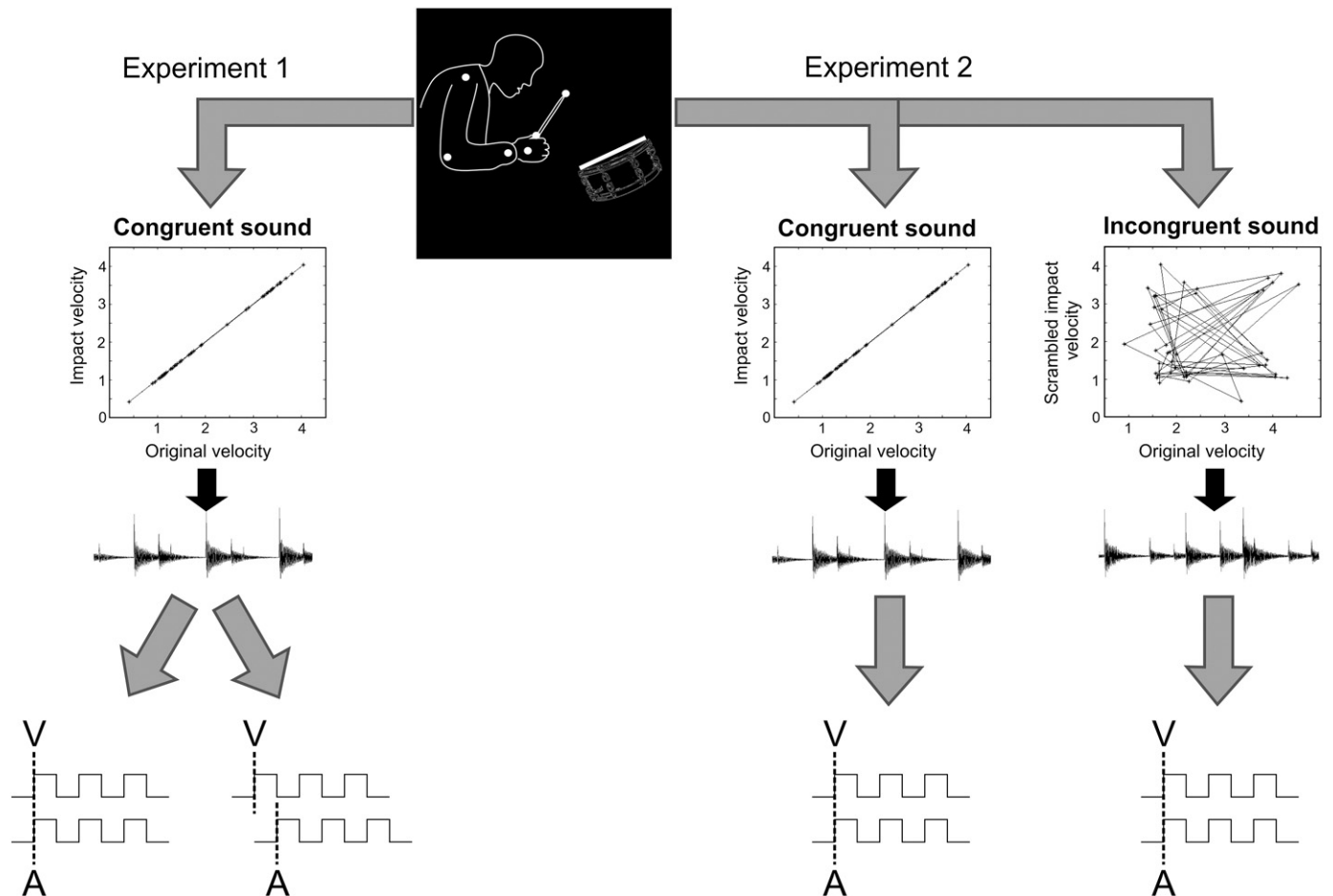
Fig. 1. Schematic of stimulus conditions in Experiments 1 and 2. Left column shows stimulus attributes in Experiment 1, whereas the two right columns depict stimulus attributes in Experiment 2. In the top center of the figure a single frame from the point-light display is presented. The point-light dots represent the drummer's arm beginning at the shoulder joint. Note that the white line outlining the drummer is presented here for clarity only and did not appear in the presented stimulus. For both experiments in left and right columns, the attributes of the visual motion, in terms of the relationship of the original motion velocity relative to implied velocity, appear in the top plots, and the produced sound waveforms appear directly under that. The lowermost panels depict the relationship of the timing of auditory (A) and visual (V) stimuli relative to one another. In Experiment 1 (left column), the displays had an audio that maintained the natural covariation with the visual signal but was presented either in synchrony (left plot on the bottom) or asynchrony (right plot on the bottom). In Experiment 2, the displays had an audio that was always in synchrony with the visual signals, although in one case it covaried with it (left plot on the top) and in the other case it did not (right plot on the top).

the natural covariation between the visually apparent velocity and the sound intensity was removed.

Audiovisual displays

The 60 Hz movies (AVI) and audio (WAV) files were imported into Adobe Premiere 1.5, where they were combined and synchronized to produce zero-lag audiovisual displays. Audiovisual displays with asynchronous audio and video were generated by either delaying video with respect to audio by 4, 8, 12 and 16 frames (corresponding to lags: −66.67, −133.33, −200, and −266.67 ms) or by delaying audio with respect to video by 4, 8, 12 and 16 frames (corresponding to lags: 66.67, 133.33, 200, and 266.67 ms). This yielded a total of nine audiovisual display files, including the one display with no delay.

To select the particular part of the audio and visual streams to use as stimuli, the initial five seconds of both files were first discarded to ensure the stabilization of the drummer performance. Next, a sound selection of nine impacts was made, always starting from two frames before the first impact and ending at one frame before the 10th impact. The sound selection was kept constant for the nine different lags, while the video was selected for each time accordingly. The resulting audiovisual files each had a duration of three seconds. Here it should be noted that since the offsets were relatively short compared with the entire duration available, there were always

both audio and visual signals present at each instant of the three-second displays. The resulting QuickTime movies were finally compressed using QuickTime Pro 6. (An example of the original zero-delay movie is provided in the Supplementary materials: Movie1.mov.).

For the behavioral study, the audiovisual files were shown to participants by using Showtime (Watson and Hu, 1999), a component of the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) extension to Matlab on a Macintosh PowerBook G3 running OS9. For the fMRI study, Presentation V11 (Neurobehavioral Systems, CA) was used to present stimuli, to log behavioral responses, and to record the times when MRI volumes were sampled with respect to stimulus delivery.

Pre-fMRI experiment psychophysical study

Here, we describe the two initial test sessions completed by each participant; these methods were adapted from closely related research (Petrini et al., 2009a,b, 2010). Observers sat in a quiet, darkened room at a distance of approximately 100 cm from a computer monitor, and wore Beyer Dynamic DT Headphones. The maximum extent of the drumming figure was 5.6 cm in the horizontal direction, resulting in a visual angle of 3.2° for the visual displays.

The experiment consisted of two blocks of 90 stimuli, run with a short rest in between. Each block contained 90 stimuli, consisting of a

random ordering of the two audiovisual conditions (natural covariation maintained, natural covariation eliminated) with nine time lags ($-266.67, -200, -133.33, -66.67, 0, 66.67, 133.33, 200, 266.67$ ms) and five repetitions of each item. Participants were instructed to press '1' on the keypad if the drummer's movements were perceived to be in synchrony with the sound, or press '3' if they were perceived as being asynchronous. After three training trials, the experimenter left the participant alone to perform the experiment.

The behavioral experiment produced 90 synchronization responses for both the displays presenting the natural covariation and the displays presenting the unnatural covariation, which were distributed as 10 repetitions at each of the nine audiovisual timings — from sound preceding video by 266 ms, to sound lagging video by 266 ms. The data were plotted as numbers of synchrony responses for each of the nine timings, and this data was fit with a Gaussian function. The peak of the Gaussian fit revealed the best perceived audiovisual synchrony for both naturally covarying and unnaturally covarying stimuli for each participant, and the tail of the distribution furthest from the peak revealed the best perceived audiovisual asynchrony. The displays with the best perceived synchrony and best perceived asynchrony were selected as stimuli for Experiments 1 and 2 of the fMRI study as appropriate.

fMRI experiments: stimulus delivery

A Windows PC (Dell Precision 690) running Presentation V11.1 (Neurobehavioral Systems, CA) presented stimuli and logged behavioral responses and MRI data acquisition pulses for each acquired brain volume. The first MRI scanner pulse initiated the presentation script, which presented an initial eight-second cue to the participants that the imaging run was beginning, during which a total of four MRI excitations were performed without MRI data acquisition to achieve steady-state magnetization. Visual clips were projected through the MRI scanner's control room window onto a screen at the participant's feet using a video projector (NEC Corporation, LT10 DLP). The participant lay supine in the MRI scanner and viewed the display through a mirror mounted on the quadrature head coil. The stimulus size was adjusted so that the visual angle of the drummer was identical to the 3.2° of visual arc used in the initial behavioral experiments. Audio was presented through a high-quality sound card interface (CDX01, Digital Audio). A sound mixer (1642VLZ pro mixer, Mackie Inc.) and commercially available MR-compatible electrostatic ear buds (STAX SRS-005 Earspeaker system; Stax LTD., Gardena, CA) worn under sound-attenuating ear muffs were used to deliver sound to the participant's ear.

In both experiments, a trial began with the three-second audiovisual clip, and participants responded with a button press before video clip offset. After offset the trial ended with a blank (black) screen. The blank screen was presented for an average of four seconds, with this duration varying randomly between two seconds and six seconds. A total of 50 trials were presented in randomized order during an imaging run which lasted 6 min and 40 s. Experiment 1 contained three runs and Experiment 2 contained two runs.

In fMRI Experiment 1, there were three stimulus conditions: Synchronous (*Matching*), Asynchronous (*Mismatching*), and a null condition containing just fixation. Participants indicated with a two-button forced choice response whether sound and video were appropriately matched or mismatched. The synchronous and asynchronous items were obtained from the pre-scan psychophysical study.

In fMRI Experiment 2, there were two stimulus conditions: naturally covarying (matching), unnaturally covarying (mismatching). We did not have a null condition in Experiment 2 because we needed to strike a balance between reducing the overall duration of the experiment and including an optimal number of conditions. Indeed, if the activation differences of Experiments 1 and 2 had

overlapped, then the additional null condition in Experiment 1 would have allowed us to carry out the analysis for this second experiment within the region of interest (ROI) obtained from Experiment 1. If, on the other hand, such overlap did not occur, we would have been able to eliminate the null condition from the main analysis of Experiment 1 in order to carry out the same mixed factorial analysis for both experiments. Participants indicated with a two-button choice response whether sound and video were appropriately matched or mismatched. Since the display with the best perceived synchrony for the naturally covarying condition had been used in Experiment 1 with the same participants, we created a new display with the same timing but taken from a non-overlapping section of the original motion capture of the drumming performance.

fMRI data acquisition

Functional images

During each fMRI experiment for each participant, we acquired 200 volumes per run of blood oxygen-level dependent (BOLD) activity in a near-whole brain acquisition, using a gradient echo spiral in-out sequence (36) on a 3 Tesla Horizon HD MRI scanner (General Electric Medical Systems, Inc.). A total of 600 volumes were acquired for Experiment 1, whereas 400 volumes were acquired for Experiment 2. Functional images used in subsequent analyses consisted of averaged images of spiral-in and spiral-out trajectories, reconstructed offline using routines written in C (courtesy of G. Glover, Stanford University, CA) and running under Linux (Fedora Core Release 5, Raleigh, NC). The averaged spiral in-out trajectories optimized sampling from brain regions prone to susceptibility artifacts and MR signal drop-out. A total of 22 axial slices beginning from the vertex (4 mm thick with 1 mm gap) were acquired with the following parameters: matrix = 128×128, FOV = 240 mm, (in plane resolution = 1.875 mm), bandwidth = 125, and TE/TR = 35/2000 ms.

Anatomical images

For each participant, we acquired a near whole-brain T1-weighted anatomical volume with identical slice prescription to functional images, and a whole-brain high-resolution anatomical SPGR volume (1.5 mm × 0.9375 mm × 0.9375 mm, FOV = 240, matrix 256 × 256, 124 slices).

fMRI data analysis

The functional and anatomical images were analyzed using Brain Voyager QX 1.9.10 (Brain Innovation, Maastricht, The Netherlands).

Pre-processing of functional data

Functional imaging data (ANALYZE format) were pre-processed by performing a slice scan time correction. Slice scan time correction was performed using sinc interpolation based on information about the TR (2000 ms) and the order of slice scanning (ascending, interleaved). 3D motion correction (6 df) was performed to detect and correct for small head movements by spatial alignment of all volumes of a subject to the first volume by rigid body transformations. Estimated translation and rotation parameters were inspected and never exceeded 3 mm or 2°. A linear trend removal and temporal high-pass filtering were then applied to remove low-frequency non-linear drifts of three or fewer cycles (0.0063 Hz) per time course. The functional MR images were smoothed using a Gaussian filter with full-width at half-maximum (FWHM) equal to 8 mm.

Pre-processing of anatomical data

The anatomical data (ANALYZE format) of each subject was loaded and converted into BrainVoyager's internal 'VMR' data format. The data were then resampled to 1 mm resolution and transformed into anterior commissure–posterior commissure (AC–PC) and Talairach

standard space. The three spatial transformations were combined and applied backward in one step to avoid quality loss due to successive data sampling. The two affine transformations, iso-voxel scaling and AC–PC transformation, were concatenated to form a single 4×4 transformation matrix. For each voxel coordinates in the target (Talairach) space a piecewise affine 'Un-Talairach' step was performed, followed by application of the inverted spatial transformation of the aforementioned matrix. The computed coordinates were used to sample the data points in the original 3D space using sinc interpolation.

Normalization of functional data

To transform the functional data into Talairach space, the functional time series data of each subject was first coregistered with the subject's 3D anatomical dataset, followed by the application of the same transformation steps as performed for the 3D anatomical dataset (see above). This step results in normalized 4D volume time course ('VTC') data. In order to avoid quality loss due to successive data sampling, normalization was performed in a single step combining a functional–anatomical affine transformation matrix, a rigid-body AC–PC transformation matrix, and a piecewise affine Talairach grid scaling step. As described for the anatomical normalization procedure, these steps were performed backward, starting with a voxel in Talairach space and sampling the corresponding data in the original functional space. The functional slices were coregistered to the anatomical volume using manual alignment to obtain optimal fit and reduce as much as possible the geometrical distortions of the echo-planar images. The necessary scaling adjustment was done interactively, using appropriate transformation and visualization tools of BrainVoyager QX.

Analysis

First level analysis

Analyses were performed on the data of individual participants using multiple linear regression of the BOLD-response time course in each voxel using three predictors (Matching, Mismatching and Blank) in Experiment 1, and two predictors (Matching, Mismatching) in Experiment 2. For each run of each participant's event-related data, a BrainVoyager protocol file (PRT) was derived, representing the onset and duration of the events for the different conditions. Predictors' time courses were adjusted for the hemodynamic response delay by convolution with a hemodynamic response function.

Second level analysis

Statistical evaluation of group data was based on a second-level GLM random effects analysis. For Experiment 1 we carried out a 2 (expertise: drummers and novices) $\times$ 2 (audiovisual synchrony: matching and mismatching) analysis of variance, with expertise as the between-participants factor and audiovisual synchrony as the within-participants factor. Similarly, for Experiment 2 we carried out a 2 (expertise: drummers and novices) $\times$ 2 (audiovisual covariation: matching and mismatching) analysis of variance, with expertise as the between-participants factor and audiovisual covariation as the within-participants factor. In both experiments the statistical maps were corrected for multiple comparisons using cluster-size thresholding (Forman et al., 1995; Goebel et al., 2006). In this method, for each statistical map the uncorrected voxel-level threshold was set at $P < 0.001$, and then was submitted to a whole-brain correction criterion based on the estimate of the map's spatial smoothness and on an iterative procedure (Monte Carlo simulation) for estimating cluster-level false-positive rates (i.e. the theoretical number of "false" positive voxels that are activated in each random map). After 1000 iterations the minimum cluster-size that yielded a cluster-level false-positive rate of 5% was used to threshold the statistical map. The

minimum cluster-size for $P < 0.05$ is reported according to the original table (in voxels) and the interpolated table (in mm³).

Results

In the Results and Discussion sections we focus on activation differences between experts and novices, as drummers are more experienced with matching audiovisual drumming actions.

Experiment 1

Behavioral data acquired during the fMRI scanning session

In Experiment 1, behavioral data acquired during MRI scanning sessions indicate that both novices and experienced drummers classified the audiovisual synchrony and asynchrony displays above 90% correctly. Comparing performance between novices and drummers failed to show significant differences between the two groups' abilities to discriminate the synchronous display as matching and the asynchronous one as mismatching (independent samples *t*-test: $t(20) = 0.167$, $p = 0.869$, two-tailed), and this is not surprising given that each participant was tested with individualized stimuli based on previous psychophysical testing. This is important, as differences in brain activity between the groups *cannot* be attributed to novices finding the task more difficult than drummers. As well as comparing the classification performances, we also compared the response time of the two groups. To this end we carried out a 2 (expertise: drummers, novices) $\times$ 2 (stimuli: matching, mismatching) $\times$ 3 (experimental run) mixed model ANOVA, with expertise as the between-subjects factor and stimuli and run as within-subjects factors. This analysis showed that only the main factor 'stimuli' ($F(1, 20) = 6.694$, $p = 0.018$) and the interaction between 'stimuli' and 'expertise' ($F(1, 20) = 13.085$, $p = 0.002$) affected significantly the response time (Fig. 2, top panels). Paired samples *t*-test, corrected for multiple comparisons, showed that whilst drummers differed in the response time for matching and mismatching displays ($t(10) = 3.756$, $p = 0.008$), novices did
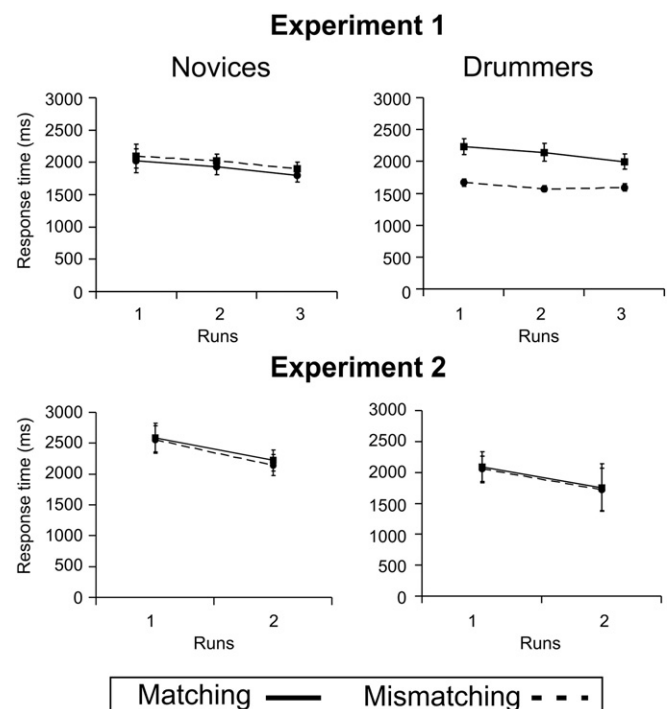

Fig. 2. The response time of drummers (right panels) and novices (left panels) is displayed for Experiment 1 (top panels) and Experiment 2 (bottom panels) as a function of experimental runs.
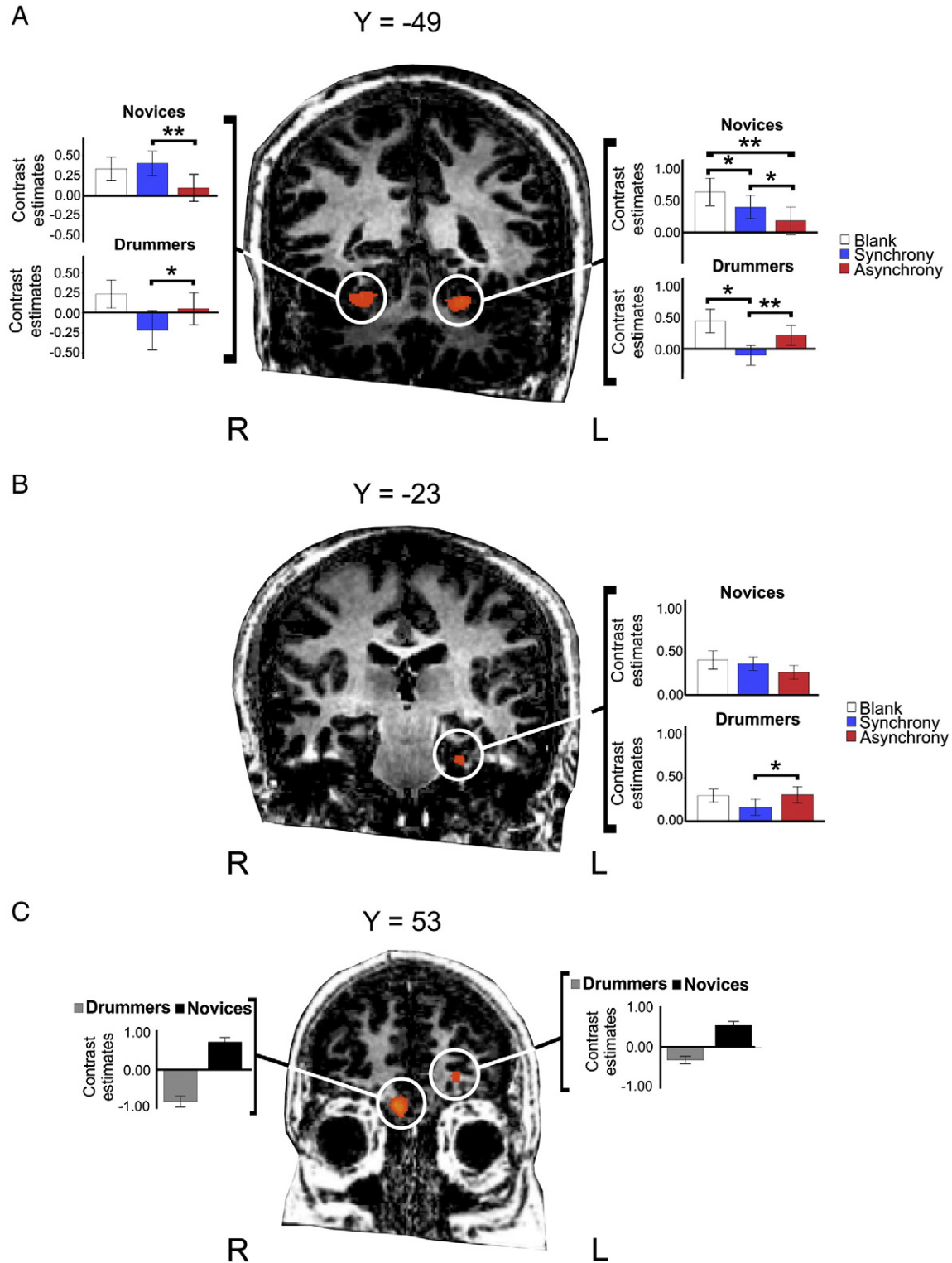
Fig. 3. Experiment 1 fMRI group activation data: Clusters of activation for which the difference between the brain responses to the two types of stimuli (audiovisual matching and mismatching drumming actions) varied across the two groups of participants (drummers and novices). Coronal slices show activation foci at two Talairach y co-ordinates. The average contrast estimates (beta weights) and relative standard errors are shown in separate histograms for drummers and novices and in different colors for matching (blue) and mismatching (red) conditions. Average contrast estimates for the blank condition are also reported in white. A) Bilateral cerebellum; B) left parahippocampal gyrus; C) clusters of activation for which the brain responses varied across the two groups of participants (drummers and novices) irrespective of the display type. The coronal slice shows activation foci at one Talairach y co-ordinate. The average contrast estimates (beta weights) and relative standard errors for drummers (gray) and novices (black) are shown in separate histograms for the right and left hemispheres. Note that in Experiment 1, mismatch between the drummer's movements and the sound reflects temporal asynchrony. * = p<0.05; ** = p<0.01 (after applying a Bonferroni's correction for three comparisons).

not ($t(10) = -0.914$, $p = 0.382$). Specifically, drummers were found to be slower when responding to matching displays than mismatching displays (Fig. 2, right top panel).

fMRI activation data

A two-way ANOVA[1] (participant group × audiovisual synchrony) revealed a main effect of expertise bilaterally in the MFG (middle frontal gyrus), no main effect of synchrony, and a significant interaction between expertise and synchrony bilaterally in the cerebellum (Fig. 3A) and left parahippocampal gyrus (Fig. 3B). Examination of the contrast parameters in these areas for the three left-handed and eight right-handed novices indicated that handedness did not affect the found differences or lack of such between matching and mismatching stimuli within this group (Supplemental Fig. 1).

The anatomical location and details of the activated foci are listed in Table 1. To control for multiple comparisons, each map was adjusted to an initial P value of <0.001 (uncorrected) and then submitted to a volume-based cluster-threshold algorithm yielding a new map thresholded at P<0.05 (corrected) with a minimum cluster threshold. The minimum cluster threshold that yielded a cluster-level false-positive rate of 5% was k = 4, 80 mm³ for the map assessed for expertise × audiovisual synchrony, and k = 3, 81 mm³ for the map assessed for expertise (Goebel et al., 2006). Average activation (i.e. averaged beta weights) indicated that the main effect of expertise was attributable to a lower level of activation in drummers relative to novices in the left and right MFG. An interaction effect in the bilateral cerebellum (Figs. 3A and 4A) and in the parahippocampal gyrus (Fig. 3B) was also attributable to reduction in activity in these areas for drummers, but not for novices, when viewing synchronous audiovisual displays. Activity in these areas during the blank condition (Fig. 3) suggested that the significant difference in activation was produced by negative BOLD responses for drummers. For example, Fig. 3 shows a reduction in activity in the bilateral cerebellum and parahippocampal gyrus when drummers viewed audiovisual matching displays (in blue) relative to blank displays (in white).

Fig. 4A shows how the reduction in activity we found bilaterally in the cerebellum, when going from an average of zero years of drumming experience for the novices to an average of 24 years for the drummers, overlaps with the reduction in activity found by Lehericy et al. (2005) in their motor sequence learning study. A regression analysis confirmed this observation, indicating that the brain activity in the cerebellum negatively correlated with the difference between novices' and drummers' years of experience (right cerebellum: $r = -0.436$, $p = 0.042$; left cerebellum: $r = -0.413$, $p = 0.056$). We did not find, however, any significant correlation between the brain activity (betas weights) for each experimental condition (matching, mismatching) and the years of experience within our drummers group in either the right (matching: $r = 0.052$, $p = 0.880$; mismatching: $r = 0.028$, $p = 0.935$) or left cerebellum (matching: $r = -0.121$, $p = 0.724$; mismatching: $r = -0.195$, $p = 0.565$). Neither was a significant correlation found between brain activity and drummers' experience in the left parahippocampal gyrus (matching: $r = -0.073$, $p = 0.830$; mismatching: $r = 0.071$, $p = 0.835$). This further analysis suggests that there was no further reduction in activity in these areas

---

[1] We did not include the blank condition in the ANOVA to be consistent with the analysis carried out in Experiment 2, thus allowing an easier comparison of the results. Nevertheless, including the blank condition and carrying out a 2 (participants: novices, drummers) × 3 (stimulus condition: matching, mismatching, and blank) repeated measure ANOVA with expertise as the between factor and stimulus condition as the within factor did not change the main findings. For the right cerebellum ($F(2, 19) = 14.492$, $p<0.001$), the left cerebellum ($F(2, 19) = 10.342$, $p = 0.001$), and the left parahippocampal gyrus ($F(2, 19) = 5.465$, $p = 0.013$) the higher order interaction was still significant. The same holds true for the right MFG ($F(1, 20) = 29.462$, $p<0.001$), and the left MFG ($F(1, 20) = 18.033$, $p<0.001$), for which the between factor was still significant.

Table 1
Experiment 1 clusters of activation from a 2×2 ANOVA with 'expertise' as a between-participants factor and 'audiovisual synchrony' as a within-participants factor. Legend: BA — Brodmann's area; MFG = middle frontal gyrus.

Anatomical region	Hemisphere	Talairach coordinate (x,y,z)	Number of voxels	Effect size[a] $F(1, 20)$	P	BA
Expertise (novices>experts)						
MFG	Right	10, 54, −12	539	18.77	0.0004	10
MFG	Left	−19, 54, 3	108	16.29	0.0006	10
Expertise ×Synchrony						
Cerebellum	Right	30, −43, −25	1442	16.35	0.0006	
Cerebellum	Left	−17, −58, −21	1608	17.83	0.0005	
Parahippocampal gyrus	Left	−26, −23, −29	191	16.95	0.0005	36

[a] Effect size = average F value for all voxels in the ROI.

after 13 years of drumming experience (i.e. the bottom range of years of experience in our drummers group).

Since we found that drummers were slower than novices (Fig. 2) when they had to respond to matching displays than mismatching displays, we further investigated whether this correlated with the contrast parameters in the regions of interest assessed in Experiment 1. A regression analysis showed a significant negative correlation between the contrast parameters obtained from the left cerebellum and the response time ($r = -0.480$, $p = 0.024$; Fig. 5, top left panel), as well as between the contrast parameters obtained from the right cerebellum and the response time ($r = -0.460$, $p = 0.031$; Fig. 5, top right panel). No significant correlation was found between the contrast parameters of right and left cerebellum and response time for the mismatching displays (right: $r = -0.158$, $p = 0.482$; left: $r = -0.406$, $p = 0.061$; Fig. 5, bottom panels), and between the contrast parameters of the left parahippocampal gyrus and the response time for the matching ($r = -0.183$, $p = 0.416$) and mismatching displays ($r = -0.253$, $p = 0.257$). The top panels of Fig. 5 show that the differences between novices and experts contribute to this regression. Indeed, the lower brain activity of the drummers group in the bilateral cerebellum corresponded to a longer response time to the matching displays; in contrast, the greater brain activity of the novices group in the bilateral cerebellum corresponded to a shorter response time to the matching displays.

Experiment 2

Behavioral data acquired during the fMRI scanning session

In Experiment 2, participants' judgments of whether or not the sight and sound matched showed a significant difference (independent samples t-test: $t(20) = 3.902$, $p = 0.001$, two-tailed) between the drummers (80% correct judgments) and novices (50% correct judgments), with the novices not being significantly different from chance (one sample t-test: $t(10) = 0.111$, mean = 50.54, two-tailed, Standard Error = 4.90). Similarly to Experiment 1, we also compared the response time of the two groups. To this end we carried out a 2 (expertise: drummers, novices) × 2 (stimuli: matching, mismatching) × 2 (experimental run) mixed model ANOVA with expertise as the between-subjects factor and stimuli and run as within-subjects factors. This analysis showed that only the main factor 'run' ($F(1, 20) = 10.680$, $p = 0.004$) affected significantly the response time, because there was a decrease in response time when going from run 1 to run 2 (Fig. 2, bottom panels).

fMRI activation data

A two-way ANOVA (participant group × audiovisual covariation) revealed no main effect of expertise. However, an extended network
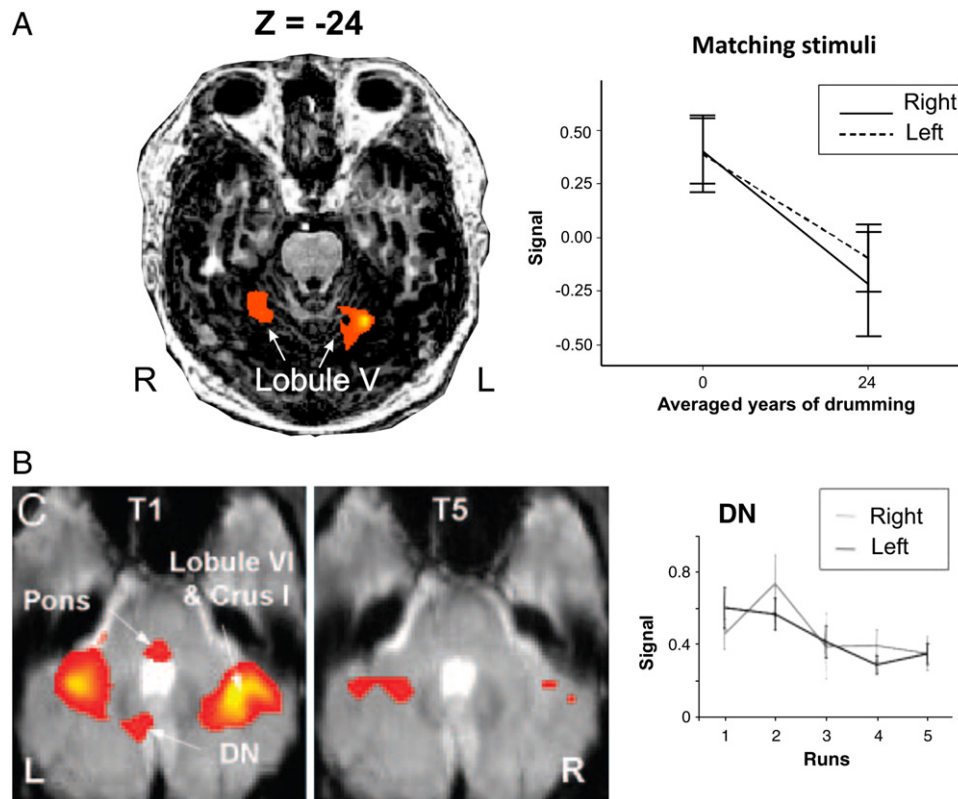
Fig. 4. A) Axial slice show the bilateral activation in lobule V of the cerebellum at one Talairach z coordinate. The average contrast estimates (beta weights) and relative standard errors for the matching condition are shown as a function of average years of drumming experience on the right; B) axial slice showing the bilateral activation in lobules V and VI as a function of days of motor training, as described by Lehericy et al. (2005). T1 = day 1; T5 = day 28. (With kind permission from the National Academy of Sciences, PNAS, Lehericy et al. (2005) Distinct basal ganglia territories are engaged in early and advanced motor sequence learning. Proc. Natl. Acad. Sci. U. S. A. 102: 12566–12571, Fig. 1C, Copyright (2005) National Academy of Sciences, U.S.A.).
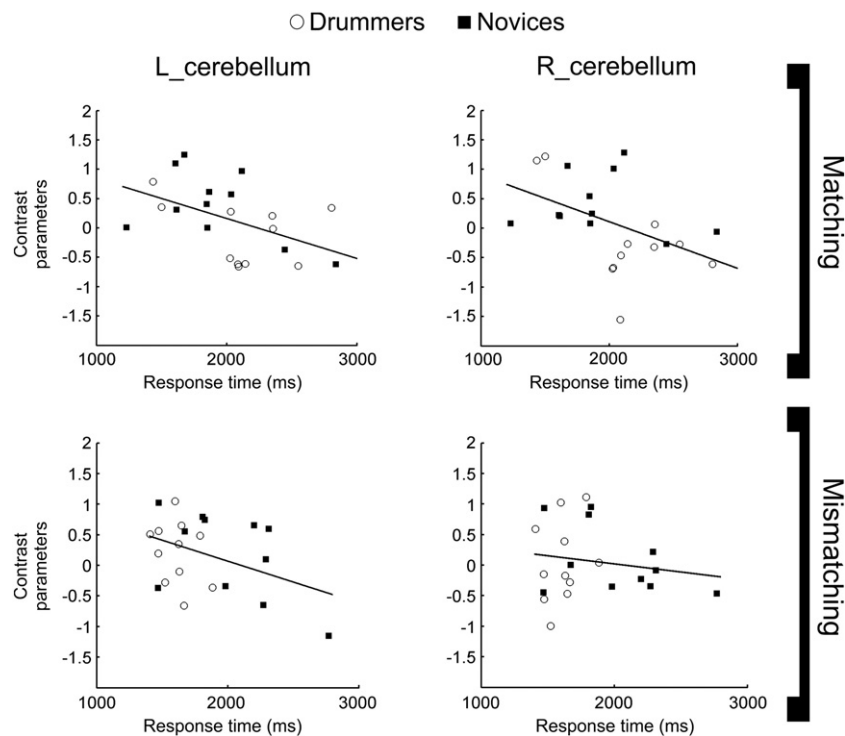

Fig. 5. The linear relation between the participants' contrast estimates (beta weights) and response time is displayed for the right (right panels) and left cerebellum (left panels) and separately for the matching (top panels) and mismatching displays (bottom panels).
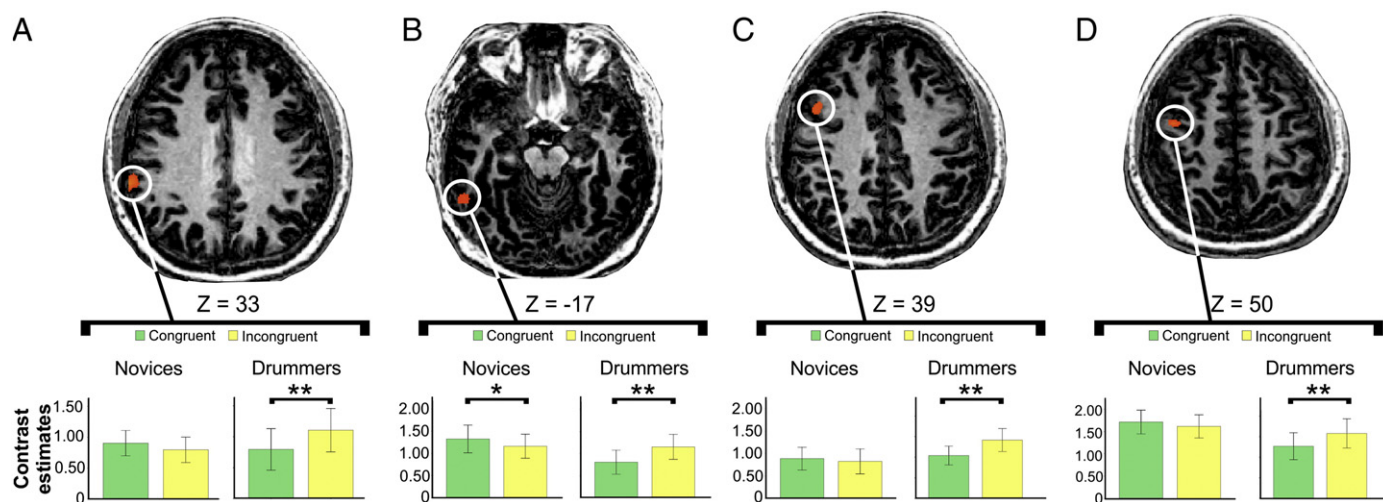
Fig. 6. Experiment 2 fMRI group activation data: Clusters of activation for which the difference between the brain responses to the two types of stimuli (matched and mismatched drumming actions) varied across the two groups of participants (drummers and novices). Axial slices show activation foci at four Talairach z coordinates. The average contrast estimates (beta weights) and relative standard errors are shown in separate diagrams for drummers and novices and in different colors for the matching (green) and mismatching (yellow) conditions. A) Right inferior parietal lobule; B) right inferior temporal gyrus; C) right middle frontal gyrus; D) right precentral gyrus. Note that in Experiment 2 the mismatch between the drummer's movements and sound reflects the lack of correspondence between strike velocities and sound intensities, while the temporal synchrony between the signals is maintained. $^{*} = p < 0.05$; $^{**} = p < 0.01$.

of significant areas for the main effect of audiovisual covariation, and a significant interaction between expertise and covariation in the right IPL (inferior parietal lobule: Fig. 6A), right ITG (inferior temporal gyrus: Fig. 6B), right MFG (Fig. 6C), and right precentral gyrus (Fig. 6D) were observed. Examination of the contrast parameters in these areas for the three left-handed and eight right-handed novices indicated that handedness did not affect the found differences or lack of such between matching and mismatching stimuli within this group (Supplemental Fig. 2). The anatomical location and details of the activated foci are listed in Table 2. To control for multiple comparisons, each map was adjusted to an initial P value of <0.001

Table 2

Experiment 2 clusters of activation from a 2×2 ANOVA with 'expertise' as a between-participants factor and 'audiovisual covariation' as a within-participants factor. Legend: BA — Brodmann's area; MFG = middle frontal gyrus; MOG = middle occipital gyrus; IFG = inferior frontal gyrus; SFG = superior frontal gyrus; STG = superior temporal gyrus; IPL = inferior parietal lobule; ITG = inferior temporal gyrus.

Anatomical region	Hemisphere	Talairach coordinate (x,y,z)	Number of voxels	Effect size[a] $F(1, 20)$	P	BA
Audiovisual covariation (matching > mismatching)						
MOG	Right	45, −76, 3	2294	18.69	0.0005	19
Cerebellum	Right	33, −71, −19	459	17.27	0.0005	
Lingual Gyrus	Right	34, −60, 9	849	16.56	0.0006	19
Cerebellum	Right	31, −53, −20	462	17.35	0.0005	
Cerebellum	Right	35, −36, −21	337	16.92	0.0006	
Posterior Cingulate	Left	−3, −62, 14	7084	19.43	0.0004	23
Cerebellum	Left	−23, −72, −24	1985	17.21	0.0005	
Audiovisual covariation (mismatching > matching)						
Insula	Left	−43, 13, 3	6083	20.52	0.0003	13
STG	Left	−53, −43, 16	1985	19.12	0.0004	13
IFG	Right	48, 21, 18	3695	18.01	0.0004	45
IFG	Right	34, 21, −9	1047	21.59	0.0003	47
SFG	Left	0, 16, 51	4802	21.50	0.0002	8
Expertise × Covariation						
IPL	Right	67, −33, 33	496	19.94	0.0003	40
ITG	Right	60, −50, −13	633	16.37	0.0006	20
MFG	Right	49, 11, 39	334	16.64	0.0006	8
Precentral Gyrus	Right	44, −4, 50	275	18.24	0.0004	6

[a] Effect size = average F value for all voxels in the ROI.

(uncorrected) and then submitted to a volume-based cluster-threshold algorithm yielding a new map thresholded at P<0.05 (corrected) with a minimum cluster threshold. The minimum cluster threshold that yielded a cluster-level false-positive rate of 5% was k=3, 77 mm^3 for the map assessed for expertise × audiovisual covariation, and k=4, 82 mm^3 for the map assessed for audiovisual covariation (Goebel et al., 2006).

The effect of interaction, similarly to Experiment 1, was due to a reduction in activity in these areas for the drummers, but not for the novices, when viewing the audiovisual matching displays (Fig. 4 in green). We did not find any significant correlation between the brain activity (betas weights) for each experimental conditions (matching, mismatching) and the years of drumming experience in either right IPL, right ITG, right MFG, or right precentral gyrus.

Discussion

Here we used novel audiovisual biological motion stimuli to investigate how action expertise modulates processing of matching and mismatching audiovisual actions. Although the effect of action expertise has been previously studied by using dance (Calvo-Merino et al., 2005; 2006) and music performance (Haslinger et al., 2005; Hodges et al., 2005; Musacchia et al., 2007; Lahav et al., 2007; Vogt et al., 2007), the way perceptual and motor expertise alters audiovisual processes is still poorly understood. Thus in Experiments 1 and 2 we examined the role of perceptual and motor expertise when processing audiovisual human actions.

Experiment 1

The results of Experiment 1 show that the brain's responses to audiovisually matching (synchronous) and mismatching (asynchronous) stimuli varied with action expertise. Specifically, we found a reduction of activity in the drummers group when viewing matching displays bilaterally in the lobule V of the cerebellum and left para-hippocampal gyrus. A reduction in neural activity in response to audiovisual training has recently been reported (Powers et al., 2010) and can be related to their earlier findings of increased sensitivity to audiovisual asynchrony (Powers et al., 2009). Since in our previous behavioral studies we repeatedly found that drummers were much more sensitive to audiovisual asynchrony than novices (Petrini et al.,

2009a,b, 2010), we believe that the reduction in brain activation for drummers when viewing over-learned drumming actions may reflect this acquired higher sensitivity.

Cerebellum

The bilateral activation of the cerebellum was mainly found at the level of the primary fissure (hemisphere lobule V; see Schmahmann et al., 1999, for a detailed atlas of the cerebellum) and was caused by *a reduction in activity* in drummers in these areas when viewing the matched audiovisual displays. An area very similar in peak coordinates (Talairach: $x = -20$, $y = -56$, $z = -28$; $x = -16$, $y = -54$, $z = -20$) to our left cerebellum was found in a PET study by Sadato et al. (1997) when comparing either tapping movements or bimanual abduction–adduction movements of the index finger to the rest condition (where no movements were performed by participants). A more recent fMRI study by Lehericy et al. (2005) showed that a motor sequence learning task decreased brain activation bilaterally in lobules V and VI of the cerebellum as movement learning progressed (Flament et al., 1996; Doyon et al., 2002; Imamizu et al., 2000). Finally, Buccino et al. (2004) reported bilateral activation of the cerebellum, at very similar Talairach coordinates to ours, when musical novices imitated observed chords on guitar frets, and when they played chords of their choice. Our results thus agree with previous findings implicating lobule V of the cerebellum as one of the areas pertaining to a brain network that subtends action-sound representation and learning of new patterns of movement (Lehericy et al., 2005; Buccino et al., 2004). Interestingly, drummers were found to be slower than novices when judging the matching displays, and this behavioral measure was found to correlate with the difference in brain activity shown by the two groups. Thus, a reduction in brain activity of the cerebellum might affect the timing of the motor response. An alternative explanation by which drummers were slower because they found the task more difficult than novices is not plausible, as demonstrated by the behavioral data during the scan and also by the behavioral data obtained from previous behavioral studies (e.g. Petrini et al., 2009a).

Additionally, our results support the idea of the cerebellum being important for controlling motor timing (Chen et al., 2006; Zatorre et al., 2007) and for computing predictive models of movement that would include movement timing (Bastian, 2006; Ohyama et al., 2003). Indeed, Jantzen et al. (2004) showed that an area of the right cerebellum with peak activity (Talairach: $x = 28$, $y = -43$, $z = -28$) resembling ours activated more for synchronized to sound tapping than asynchronized to sound tapping. This result is in line with what we found for the novices group, which showed higher activation in lobule V of the cerebellum when viewing matching (synchrony) displays than when viewing mismatching (asynchrony) displays.

Parahippocampal gyrus

Besides the cerebellum, a significant effect of expertise for viewing matching and mismatching displays was found in the left parahippocampal gyrus. Similarly to the cerebellum, the parahippocampal gyrus showed a *reduction in activity* in drummers when viewing matching displays, while no difference in activation was observed between matching and mismatching displays in novices. The left parahippocampal gyrus (BA 36) is known to exhibit interaction effects between motor experience and self-ratings of ability to perform dance movements (Cross et al., 2006), and when contrasting activation elicited by meaningful actions versus meaningless actions for the explicit purpose of subsequent recognition (Decety et al., 1997). Neuropsychological studies indicate that lesions of the hippocampal region impair the ability to rapidly acquire and store new information about facts and events (Squire and Knowlton, 1995; Ungerleider, 1995). These data are supported by subsequent studies showing that left parahippocampal gyrus activation correlates with the number of unrelated words recalled (Alkire et al., 1998), and

decreases in patients with Alzheimer's disease compared with controls (Rombouts et al., 2000). Thus the reduction in activity in the left parahippocampal gyrus for the drummers when presented with the matching displays might indicate a decrease in brain activity used by this group when recalling over-learned information about drumming actions.

Experiment 2

The reduced level of activation we showed in Experiment 1 for drummers when viewing matching actions was replicated in Experiment 2. However, the interaction effects indicate that the brain's responses to audiovisual matching (naturally covarying) and mismatching (unnaturally covarying) varied with action expertise in completely different areas within the right hemisphere. These areas were the IPL (BA 40), ITG (BA 20), MFG (BA 8), and precentral gyrus (BA 6). The absence of the cerebellum from the areas found in this second experiment adds further evidence to the idea that the cerebellum mainly responds to *discrepancies in temporal occurrence* of the represented action and resulting sound, while other brain regions respond to *discrepancies in temporal congruency*. In other words, since the movements of the drummers and the drumming sound were always synchronous in this second experiment, the cerebellum maintained the same level of activation for both groups when viewing the matching and mismatching displays.

An alternative explanation of why we found such different results in Experiments 1 and 2 might be that the task in Experiment 2 was more difficult and required the participation of frontal areas (e.g. right precentral gyrus and right MFG). This would explain also why in Experiment 2, but not in Experiment 1, we found two groups of activated brain areas *independent* of expertise level, one group more activated for the matching than the mismatching condition, and another more activated for the mismatching than the matching condition (Table 2). Furthermore, this idea is supported by the fact that the novices could not correctly classify matching and mismatching displays in Experiment 2 (indeed their performance was at chance level), and by the fact that drummers' ability to correctly classify the two kinds of display decreased when going from Experiment 1 to 2 (although this group accuracy remained significantly above chance level). However, this alternative explanation fails to explain why we did not also find a difference across groups in the cerebellum and parahippocampal gyrus when comparing matching and mismatching condition, if the only difference between Experiments 1 and 2 was in task difficulty.

Inferior parietal gyrus, inferior temporal gyrus, medial frontal gyrus and precentral gyrus

Similar areas to our IPL, ITG, and precentral gyrus were found to activate more for observation of meaningless versus meaningful actions in the study of Decety et al. (1997), while the MFG activated for the opposite contrast. All or some of these four areas are active in audiomotor (Bangert et al., 2006; Baumann et al., 2007; Lahav et al., 2007), visuomotor (Grèzes et al., 1999, 2003; Perani et al., 2001; Gallagher and Frith, 2004; Buccino et al., 2004) and audiovisual representation studies (Haslinger et al., 2005). Hence, we regard these regions as being crucial for action observation/listening and imitation, as well as for action expertise.

The activation we observed in the right precentral gyrus included both the premotor ventral (PMv) and premotor dorsal (PMd) cortices, if we follow the convention proposed by Rizzolatti et al. (2002) that in humans the border between the ventral and the dorsal premotor cortex lies at the level of the upper border of the frontal eye field, approximately at Talairach coordinate $Z = 50$. A similar cluster of activation at the border between PMv and PMd has been also found by Buccino et al. (2004) during observation and imitation of guitar chords. While PMv appears to be involved in direct visuomotor

transformations, PMd appears to be involved in indirect transformations (Hoshi and Tanji, 2006; Zatorre et al., 2007). Direct transformations from sound to motor actions are very relevant for music performance and have been shown to engage PMv (Bangert et al., 2006; Baumann et al., 2007; Lahav et al., 2007), indicating that this portion of premotor cortex may activate when hearing a sound for which one has an acquired model in his/her own repertoire. On the other hand, PMd may have a more indirect role in sensory-motor transformation by retrieving and integrating sensory information to plan and execute an action (Hoshi and Tanji, 2004, 2006; Zatorre et al., 2007). In our experiment, both the activation of an acquired motor program and the integration of different sensory cues linked to the represented actions may have been essential for judging correct correspondence between drummers' movements and sound information, with consequent activation of both portions of the premotor cortex. Furthermore, the idea that PMd is involved in higher order aspects of movements, such as when increasing metrical saliency (Chen et al., 2006), and when increasing rhythmic complexity of movements (Chen et al., 2008), would explain why this area was affected by expertise level only in Experiment 2, where a more complex feature of the drumming action-sound representation was manipulated. Music, much more than visual stimuli, has a remarkable ability to drive rhythmic, metrically organized motor behavior (Zatorre et al., 2007). In Experiment 1, the sound was a selection of three repeated cycles of drumming swing groove (nine impacts) with a regular accent on the second beat, whereas in Experiment 2 the sound (nine impacts with the same impact times of those in Experiment 1) did not contain any regular rhythmic pattern because the occurrence of the accented beat was randomized. The drummer's visually presented movement, in contrast, maintained always a regular swing groove rhythmic pattern with the accent on the second beat, even when desynchronized from the sound in Experiment 1 (although the starting beat and final beat were different from those of the synchronized condition). When considering the action-sound representation, the mismatching condition in Experiment 2 was metrically more complex, thus possibly requiring the involvement of PMd.

Finally, similar areas to our right PMv (BA 6) and right IPL/supramarginal gyrus (BA 40) are known to activate when contrasting observed piano playing with and without sound with a rest condition (Haslinger et al., 2005). While PMv may be important for the recovering of acquired motor programs, the inferior parietal cortex may be important for cross-modal processing and integration of audiovisual information (e.g. Calvert et al., 2000, 2001; Calvert, 2001). The fact that in Haslinger et al. (2005) the movements were always synchronized with the sound would support the participation of the IPL/supramarginal gyrus in audiovisual integration of observed actions *only* as long as the *temporal co-occurrence* of the signals is maintained, as was the case in our Experiment 2, where both matching and mismatching displays were audiovisually synchronized.

Limitations

As indicated in the Materials and methods and Discussion sections, although we did match the level of perceived synchrony and asynchrony for each individual prior to Experiment 1, we did not match the level of perceived congruency and incongruency for each individual prior to Experiment 2. This introduces some interpretational ambiguities for the results of Experiment 2, which could have been a consequence of differences in task difficulty between the two groups as much as differences in action expertise. Our design cannot disentangle this possibility from our main interpretation of Experiment 2, and further experiments will be needed to draw strong conclusions on this point. Nevertheless the consistency of findings between Experiments 1 and 2 showing a reduction in brain activity for drummers and not for novices for matching displays is very compelling, especially when considering that drummers were really

good in discriminating the matching and mismatching displays in both experiments (Experiment 1: 92% correct; Experiment 2: 80% correct). A further study limitation concerns the relatively small number of drummers and novices included. We had the potential to enroll more participants than the number we reported, but found that gender imbalance would be a problem in data interpretation. We chose to study expert subjects from one gender who had decades of experience (mean 24 years). Hence, our sample number in this homogeneous expert group was 11, to which we additionally age-matched 11 novice controls. Indeed, finding age matched novices with minimal musical experience posed an additional challenge. Future studies comparing a greater number of musicians and non-musicians would strengthen our conclusions. These interpretational limitations, however, do not diminish the novel finding reported here, that brain activation in action-sound representation areas is reduced by long-term acquired familiarity with audiovisual represented action.

Conclusion

The results from both experiments provide converging evidence that expertise with a certain audiovisual action reduces activation in brain areas crucial for action-sound representation and audiovisual integration. Because both temporal synchrony and temporal congruency of the produced sound may be easily predicted from the observed movements using acquired internal models of that action, a reduction in activation may result as a consequence of perceptual and motor expertise (McKiernan et al., 2003; Parsons et al., 2005). However, which network of areas will reduce their activation in response to over-learned and familiar audiovisual actions will depend on the level of stimulus complexity and on the temporal correspondence between the visual action and sound information. Our findings indicate that there are *two separate networks* processing the sound and sight of others' actions: one network (including cerebellum and parahippocampal gyrus) would be responsible for processing the sensory synchrony of the auditory and visual information, while another (including fronto-temporal–parietal regions) would be responsible for processing the sensory congruency of the two sensory information. This is a novel finding that underlines the complexity of functional reorganization dictated by perceptual and motor expertise.

Several neuroimaging studies have observed that musicians show lower levels of activity in motor regions than non-musicians during the performance of simple motor tasks (Hund-Georgiadis and Von Cramon, 1999; Jäncke et al., 2000; Koeneke et al., 2004), and here we show that this is also the case when musicians are only observing and not performing an audiovisual action. Even though it is still not clear what this reduction in activity entails, we believe that it may explain why musicians have been repeatedly found to outperform non-musicians in a variety of tasks that have little to do with music (Schmithorst and Holland, 2004; Brochard et al., 2004; Magne et al., 2006; Musacchia et al., 2007). Indeed, if the brain, through this reduction in activity, is able to conserve perceptual and motor resources, then these might be reallocated to other brain areas to increase the musicians' efficiency in other everyday tasks, as suggested by some studies where the reduced activity found in certain brain areas of musicians was accompanied by an increase of activity in other areas (Hund-Georgiadis and Von Cramon, 1999; Schmithorst and Holland, 2004). Connectivity studies might give a definitive answer to this point by examining whether the reduction in activity of certain areas is followed by an increase in activity of other brain areas when musicians perform musical and non-musical tasks.

Finally, we still need to understand what are the mechanisms and what is the trajectory for developing a stable and maximal reduction in activity through musical practice, as this could have practical implications for the principles of musical training. The lack of correlation between years of drumming experience and reduction in brain activity in the present study suggests that after 13 years of

drumming there is no additional reduction in activation. However, from our findings we cannot derive when a stable and maximal reduction of activity is achieved, and future studies should examine when this happens by either comparing musicians with one to 10 years of experience or by longitudinally studying the effect of music practice on musicians' performance and brain activity. These kinds of investigations will also add valuable information about the development of brain behavior during learning.

Supplementary materials related to this article can be found online at doi:10.1016/j.neuroimage.2011.03.009.

Acknowledgments

References

Alkire, M.T., Haier, R., Fallon, J.H., Cahill, L., 1998. Hippocampal, but not amygdala, activity at encoding correlates with long-term, free recall of non-emotional information. Proc. Natl Acad. Sci. 95, 14506–14510.
Bangert, M., Peschel, T., Schlaug, G., Rotte, M., Drescher, D., Hinrichs, H., Heinze, H.J., Altenmuller, E., 2006. Shared networks for auditory and motor processing in professional pianists: evidence from fMRI conjunction. Neuroimage 30, 917–926.
Bastian, A.J., 2006. Learning to predict the future: the cerebellum adapts feedforward movement control. Curr. Opin. Neurobiol. 16, 645–649.
Baumann, S., Koeneke, S., Meyer, M., Lutz, K., Jancke, L., 2005. A network for sensory-motor integration: what happens in the auditory cortex during piano playing without acoustic feedback? Ann. NY Acad. Sci. 1060, 186–188.
Baumann, S., Koeneke, S., Schmidt, C.F., Meyer, M., Lutz, K., Jancke, L., 2007. A network for audio-motor coordination in skilled pianists and nonmusicians. Brain Res. 1161, 65–78.
Brainard, D.H., 1997. The psychophysics toolbox. Spat. Vis. 10, 433–436.
Brochard, R., Dufour, A., Després, O., 2004. Effect of musical expertise on visuospatial abilities: evidence from reaction times and mental imagery. Brain Cogn. 54, 103–109.
Brooks, A., van der Zwan, R., Billard, A., Petreska, B., Clarke, S., Blanke, O., 2007. Auditory motion affects visual biological motion processing. Neuropsychologia 45, 523–530.
Buccino, G., Binkofski, F., Fink, G.R., Fadiga, L., Fogassi, L., Gallese, V., Seitz, R.J., Zilles, K., Rizzolatti, G., Freund, H.J., 2001. Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. Eur. J. Neurosci. 13, 400–404.
Buccino, G., Vogt, S., Ritzl, A., Fink, G.R., Zilles, K., Freund, H.J., Rizzolatti, G., 2004. Neural circuits underlying imitation learning of hand actions: an event-related fMRI study. Neuron 42, 323–334.
Calvert, G.A., 2001. Crossmodal processing in the human brain: insights from functional neuroimaging studies. Cereb. Cortex 11, 1110–1123.
Calvert, G.A., Campbell, R., Brammer, M.J., 2000. Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. Curr. Biol. 10, 649–657.
Calvert, G.A., Hansen, P.C., Iversen, S.D., Brammer, M.J., 2001. Detection of multisensory integration sites by application of electrophysiological criteria to the BOLD response. Neuroimage 14, 427–438.
Calvo-Merino, B., Glaser, D.E., Grèzes, J., Passingham, R.E., Haggard, P., 2005. Action observation and acquired motor skills: an fMRI study with expert dancers. Cereb. Cortex 15, 1243–1249.
Calvo-Merino, B., Grèzes, J., Glaser, D.E., Passingham, R.E., Haggard, P., 2006. Seeing or doing? Influence of visual and motor familiarity in action observation. Curr. Biol. 16, 1905–1910.
Chen, J.L., Zatorre, R.J., Penhune, V.B., 2006. Interactions between auditory and dorsal premotor cortex during synchronization to musical rhythms. Neuroimage 32, 1771–1781.
Chen, J.L., Penhune, V.B., Zatorre, R.J., 2008. Listening to musical rhythms recruits motor regions of the brain. Cereb. Cortex 18, 2844–2854.
Cross, E.S., Hamilton, A.F.D.C., Grafton, S.T., 2006. Building a motor simulation de novo: observation of dance by dancers. Neuroimage 31, 1257–1267.
Dahl, S., 2004. Playing the accent — comparing striking velocity and timing in an ostinato rhythm performed by four drummers. Acta Acustica United Acustica 90, 762–776.
Decety, J., Grèzes, J., 1999. Neural mechanisms subserving the perception of human actions. Trends Cogn. Sci. 3, 172–178.
Decety, J., Grèzes, J., Costes, N., Perani, D., Jeannerod, M., Procyk, E., Grassi, F., Fazio, F., 1997. Brain activity during observation of actions. Influence of action content and subject's strategy. Brain 120, 1763–1777.
Doyon, J., Song, A.W., Karni, A., Lalonde, F., Adams, M.M., Ungerleider, L.G., 2002. Experience-dependent changes in cerebellar contributions to motor sequence learning. Proc. Natl Acad. Sci. U.S.A. 99, 1017–1022.

Flament, D., Ellermann, J.M., Kim, S.G., Ugurbil, K., Ebner, T.J., 1996. Functional magnetic resonance imaging of cerebellar activation during the learning of a visuomotor dissociation task. Hum. Brain Mapp. 4, 210–226.
Fontana, F., Rocchesso, D., Avanzini, F., 2004. Computation of nonlinear filter networks containing delay-free paths. 7th International Conference on Digital Audio Effects (DAFX-04), pp. 113–118. Naples, Italy.
Forman, S.D., Cohen, J.D., Fitzgerald, M., Eddy, W.F., Mintun, M.A., Noll, D.C., 1995. Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a clustersize threshold. Magn. Reson. Med. 33, 636–647.
Gallagher, H.L., Frith, C.D., 2004. Dissociable neural pathways for the perception and recognition of expressive and instrumental gestures. Neuropsychologia 42, 1725–1736.
Gallese, V., Fadiga, L., Fogassi, L., Rizzolatti, G., 1996. Action recognition in the premotor cortex. Brain 119, 593–609.
Goebel, R., Esposito, F., Formisano, E., 2006. Analysis of functional image analysis contest (FIAC) data with brainvoyager QX: from single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. Hum. Brain Mapp. 27, 392–401.
Grèzes, J., Costes, N., Decety, J., 1999. The effects of learning and intention on the neural network involved in the perception of meaningless actions. Brain 122, 1875–1887.
Grèzes, J., Armony, J.L., Rowe, J., Passingham, R.E., 2003. Activations related to 'mirror' and 'canonical' neurones in the human brain: an fMRI study. Neuroimage 18, 928–937.
Haslinger, B., Erhard, P., Altenmuller, E., Schroeder, U., Boecker, H., Ceballos-Baumann, A.O., 2005. Transmodal sensorimotor networks during action observation in professional pianists. J. Cogn. Neurosci. 17, 282–293.
Hodges, D.A., Hairston, W.D., Burdette, J.H., 2005. Aspects of multisensory perception: the integration of visual and auditory information in musical experiences. Ann. NY Acad. Sci. 1060, 175–185.
Hoshi, E., Tanji, J., 2004. Functional specialization in dorsal and ventral premotor areas. Prog. Brain Res. 143, 507–511.
Hoshi, E., Tanji, J., 2006. Differential involvement of neurons in the dorsal and ventral premotor cortex during processing of visual signals for action planning. J. Neurophysiol. 95, 3596–3616.
Hund-Georgiadis, M., Von Cramon, D.Y., 1999. Motor-learning-related changes in piano players and nonmusicians revealed by functional magnetic-resonance signals. Exp. Brain Res. 125, 417–425.
Iacoboni, M., 2005. Neural mechanisms of imitation. Curr. Opin. Neurobiol. 15, 632–637.
Imamizu, H., Miyauchi, S., Tamada, T., Sasaki, Y., Takino, R., Putz, B., Yoshioka, T., Kawato, M., 2000. Human cerebellar activity reflecting an acquired internal model of a new tool. Nature 403, 192–195.
Jäncke, L., Shah, N.J., Peters, M., 2000. Cortical activations in primary and secondary motor areas for complex bimanual movements in professional pianists. Brain Res. Cogn. Brain Res. 10, 177–183.
Jantzen, K.J., Steinberg, F.L., Kelso, J.A.S., 2004. Brain networks underlying human timing behavior are influenced by prior context. Proc. Natl Acad. Sci. 101, 6815–6820.
Johansson, G., 1973. Visual perception of biological motion and model for its analysis. Percept. Psychophys. 14, 201–211.
Keysers, C., Kohler, E., Umilta, M.A., Nanetti, L., Fogassi, L., Gallese, V., 2003. Audiovisual mirror neurons and action recognition. Exp. Brain Res. 153, 628–636.
Klin, A., Lin, D., Gorrindo, P., Ramsay, G., Jones, W., 2009. Two-year-olds with autism orient to non-social contingencies rather than biological motion. Nature 459 (7244), 257–261.
Koeneke, S., Lutz, K., Wustenberg, T., Jäncke, L., 2004. Long-term training affects cerebellar processing in skilled keyboard players. Neuroreport 15, 1279–1282.
Kohler, E., Keysers, C., Umilta, M.A., Fogassi, L., Gallese, V., Rizzolatti, G., 2002. Hearing sounds, understanding actions: action representation in mirror neurons. Science 297, 846–848.
Lahav, A., Saltzman, E., Schlaug, G., 2007. Action representation of sound: audiomotor recognition network while listening to newly acquired actions. J. Neurosci. 27, 308–314.
Lehericy, S., Benali, H., Van de Moortele, P.F., Pelegrini-Isaac, M., Waechter, T., Ugerbil, K., Doyon, J., 2005. Distinct basal ganglia territories are engaged in early and advanced motor sequence learning. Proc. Natl Acad. Sci. U.S.A. 102, 12566–12571.
Magne, C., Schon, D., Besson, M., 2006. Musician children detect pitch violations in both music and language better than non-musician children: behavioral and electro-physiological approaches. J. Cogn. Neurosci. 18, 199–211.
McKiernan, K.A., Kaufman, J.N., Kucera-Thompson, J., Binder, J.R., 2003. A parametric manipulation of factors affecting task-induced deactivation in functional neuroi-maging. J. Cogn. Neurosci. 15, 394–408.
Musacchia, G., Sams, M., Skoe, E., Kraus, N., 2007. Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. Proc. Natl Acad. Sci. 104, 15894–15898.
Nishitani, N., Hari, R., 2000. Temporal dynamics of cortical representation for action. Proc. Natl Acad. Sci. U.S.A. 97, 913–918.
Ohyama, T., Nores, W.L., Murphy, M., Mauk, M.D., 2003. What the cerebellum computes. Trends Neurosci. 26, 222–227.
Oldfield, R.C., 1971. Assessment and analysis of handedness — Edinburgh inventory. Neuropsychologia 9, 97–113.
Parsons, L.M., Sergent, J., Hodges, D.A., Fox, P.T., 2005. The brain basis of piano performance. Neuropsychologia 43, 199–215.
Pelli, D.G., 1997. The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spat. Vis. 10, 437–442.
Perani, D., Fazio, F., Borghese, N.A., Tettamanti, M., Ferrari, S., Decety, J., Gilardi, M.C., 2001. Different brain correlates for watching real and virtual hand actions. Neuroimage 14, 749–758.

Petrini, K., Dahl, S., Rocchesso, D., Waadeland, D.H., Avanzini, F., Puce, A., Pollick, F., 2009a. Multisensory integration of drumming actions: musical expertise affects perceived audiovisual asynchrony. Exp. Brain Res. 198, 339–352.

Petrini, K., Russell, M., Pollick, F., 2009b. When knowing can replace seeing in audiovisual integration of actions. Cognition 110, 432–439.

Petrini, K., Holt, P.S., Pollick, F., 2010. Expertise with multisensory events eliminates the effect of biological motion rotation on audiovisual synchrony perception. J. Vis. 10 (5), 1–14.

Powers, A.R., Hillock, A.R., Wallace, M.T., 2009. Perceptual training narrows the temporal window of multisensory binding. J. Neurosci. 29, 12265–12274.

Powers, A.R., Hevey, M., Wallace, M.T., 2010. The neural correlates of multisensory perceptual training. Abstract in IMRF2010 Conference Proceedings, Liverpool, UK.

Rizzolatti, G., Fadiga, L., Gallese, V., Fogassi, L., 1996. Premotor cortex and the recognition of motor actions. Brain Res. Cogn. Brain Res. 3, 131–141.

Rizzolatti, G., Fogassi, L., Gallese, V., 2002. Motor and cognitive functions of the ventral premotor cortex. Curr. Opin. Neurobiol. 12, 149–154.

Rombouts, S.A.R.B., Barkhof, F., Veltman, D.J., Machielsen, W.C.M., Witter, M.P., Bierlaagh, M.A., Lazeron, R.H.C., Valk, J., Scheltens, P., 2000. Functional MR imaging in Alzheimer's disease during memory encoding. AJNR Am. J. Neuroradiol. 21, 1869–1875.

Sadato, N., Yonekura, Y., Waki, A., Yamada, H., Ishii, Y., 1997. Role of the supplementary motor area and the right premotor cortex in the coordination of bimanual finger movements. J. Neurosci. 17, 9667–9674.

Saygin, A.P., 2007. Superior temporal and premotor brain areas necessary for biological motion perception. Brain 130, 2452–2461.

Saygin, A.P., Wilson, S.M., Hagler Jr., D.J., Bates, E., Sereno, M.I., 2004. Point-light biological motion perception activates human premotor cortex. J. Neurosci. 24, 6181–6188.

Schmahmann, J.D., Doyon, J., McDonald, D., Holmes, C., Lavoie, K., Hurwitz, A.S., Kabani, N., Toga, A., Evans, A., Petrides, M., 1999. Three-dimensional MRI atlas of the human cerebellum in proportional stereotaxic space. Neuroimage 10, 233–260.

Schmithorst, V.J., Holland, S.K., 2004. The effect of musical training on the neural correlates of math processing: a functional magnetic resonance imaging study in humans. Neurosci. Lett. 354, 193–196.

Squire, L., Knowlton, B., 1995. Memory, hippocampus and brain systems. In: Gazzaniga, M. (Ed.), The Cognitive Neurosciences. MIT Press, Cambridge, MA, pp. 825–838.

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.

Ungerleider, L.G., 1995. Functional brain imaging studies of cortical mechanisms for memory. Science 270, 769–775.

Vogt, S., Buccino, G., Wohlschlager, A.M., Canessa, N., Shah, N.J., Zilles, K., Eickhoff, S.B., Freund, H.J., Rizzolatti, G., Fink, G.R., 2007. Prefrontal involvement in imitation learning of hand actions: effects of practice and expertise. Neuroimage 37, 1371–1383.

Waadeland, C.H., 2006. Strategies in empirical studies of swing groove. Studia Musicologica Norvegica 32, 169–191.

Watson, A.B., Hu, J., 1999. ShowTime: a QuickTime-based infrastructure for vision research displays. Perception 28 (ECVP Abstract Supplement): 45b. .

Wilson, S.M., Saygin, A.P., Sereno, M.I., Iacoboni, M., 2004. Listening to speech activates motor areas involved in speech production. Nat. Neurosci. 7, 701–702.

Zatorre, R.J., Chen, J.L., Penhune, V.B., 2007. When the brain plays music: auditory–motor interactions in music perception and production. Nat. Rev. Neurosci. 8, 547–558.