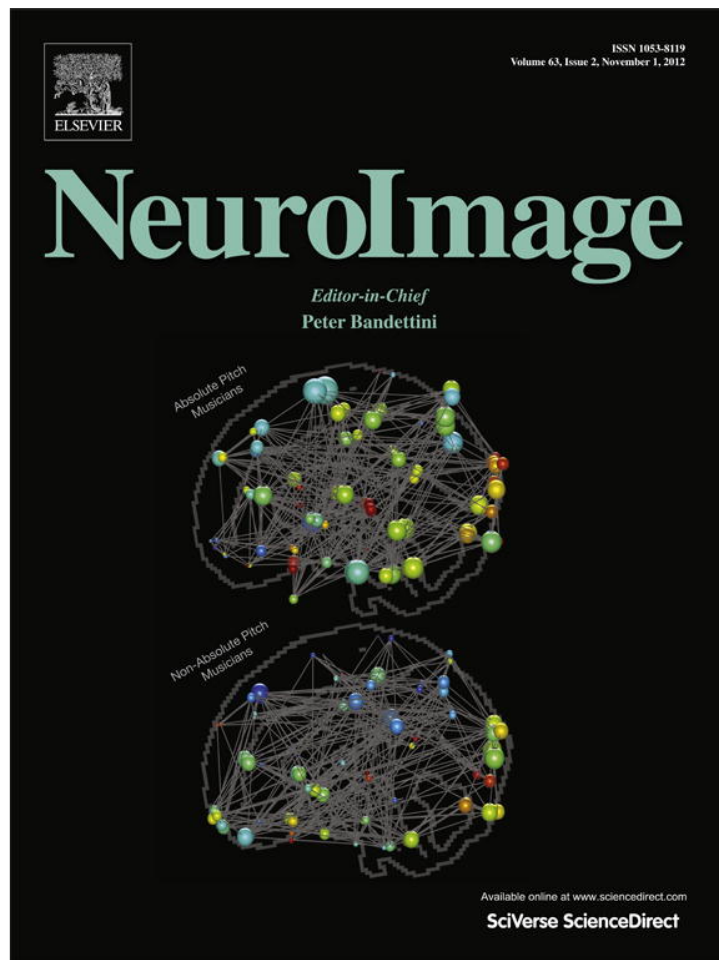


Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

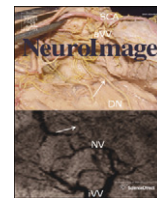
Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>

Contents lists available at [SciVerse ScienceDirect](http://www.elsevier.com/locate/ynimg)

NeuroImage

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

Learning piano melodies in visuo-motor or audio-motor training conditions and the neural correlates of their cross-modal transfer

Annerose Engel ^{a,b,c,*}, Marc Bangert ^{c,d}, David Horbank ^c, Brenda S. Hijmans ^a, Katharina Wilkens ^c, Peter E. Keller ^{c,e}, Christian Keysers ^{a,b}

^a Department of Neuroscience, University Medical Center Groningen, University of Groningen, The Netherlands

^b The Netherlands Institute for Neuroscience, Royal Netherlands Academy of Arts and Sciences, Amsterdam, The Netherlands

^c Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

^d Institute of Musician's Medicine, University of Music Carl Maria von Weber, Dresden, Germany

^e MARCS Institute, University of Western Sydney, Australia

ARTICLE INFO

Article history:

Accepted 11 March 2012

Available online 22 March 2012

Keywords:

Cross-modal transfer

Human fMRI

Piano learning

Rolandic operculum

ABSTRACT

To investigate the cross-modal transfer of movement patterns necessary to perform melodies on the piano, 22 non-musicians learned to play short sequences on a piano keyboard by 1) merely listening and replaying (vision of own fingers occluded) or 2) merely observing silent finger movements and replaying (on a silent keyboard). After training, participants recognized with above chance accuracy 1) audio-motor learned sequences upon visual presentation ($89 \pm 17\%$), and 2) visuo-motor learned sequences upon auditory presentation ($77 \pm 22\%$). The recognition rates for visual presentation significantly exceeded those for auditory presentation ($p < .05$). fMRI revealed that observing finger movements corresponding to audio-motor trained melodies is associated with stronger activation in the left rolandic operculum than observing untrained sequences. This region was also involved in silent execution of sequences, suggesting that a link to motor representations may play a role in cross-modal transfer from audio-motor training condition to visual recognition. No significant differences in brain activity were found during listening to visuo-motor trained compared to untrained melodies. Cross-modal transfer was stronger from the audio-motor training condition to visual recognition and this is discussed in relation to the fact that non-musicians are familiar with how their finger movements look (motor-to-vision transformation), but not with how they sound on a piano (motor-to-sound transformation).

© 2012 Elsevier Inc. All rights reserved.

Introduction

Imagine yourself learning to play a melody on the piano just by listening to a series of tones. While you are training to play this melody you never observe your own fingers moving — because they are occluded or shrouded in the darkness of the practice studio. Over a long run of trials (and errors) you come to learn how to use your fingers to reproduce the heard melody. Now that you can successfully play the melody, would you be later able to recognize the corresponding sequence if you observed someone else playing it silently?

This question addresses the cross-modal transfer of actions from one modality to another, or in other words, whether actions trained in connection with one domain (sound) generalize to another (vision). The assumption of such cross-modal interaction was motivated by research investigating musicians who showed cross-modal co-activations (Bangert and Altenmüller, 2003; Bangert et al., 2006; Hasegawa et al., 2004; Haslinger et al., 2005; Haueisen and Knösche, 2001; Lotze et al.,

2003). Examining both directions of audio-motor associations using functional Magnetic Resonance Imaging (fMRI), Bangert et al. (2006) asked professional pianists and non-musicians to listen passively to short piano melodies. Additionally in a different condition, participants were requested to press keys on a mute MRI-compliant piano keyboard. During listening to piano melodies, professional pianists compared to the non-musicians showed increased activity in a distributed cortical motor-related network (c.f., Haueisen and Knösche, 2001), including areas of the ventral premotor cortex, Broca's area and the inferior parietal lobe. Furthermore, executing silent finger movements on a piano keyboard elicited stronger activation of auditory-sensory areas in pianists compared to non-musicians (c.f., Lotze et al., 2003).

Exploring the development of such audio-motor associations, Bangert and Altenmüller (2003) trained non-musicians to play the piano over a period of 10 weeks. By using electroencephalography (EEG) they showed that passive piano listening leads to activity in frontal and central electrodes but is accompanied by activity in electrodes over sensori-motor areas after as little as 20 min of practice, and is even more pronounced with longer practice. They also found activation of temporal auditory areas during the execution of silent finger movements (such as playing piano) after five weeks of piano practice. Moreover, further research in

* Corresponding author at: Music Cognition & Action Group, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany.
E-mail address: engela@cbs.mpg.de (A. Engel).

the visual domain confirms that there is a strong interconnection not only between the auditory and motor but also between audio-visuo-motor domains in musicians. Hasegawa et al. (2004) found that trained musicians, but not non-musicians or less trained musicians, show activity of auditory areas, such as the planum temporale (additional to co-activation of motor regions), when observing sequences of silent key presses on a piano keyboard (c.f., Haslinger et al., 2005). Taken together, these studies reveal bidirectional cross-modal transfer between the motor and auditory domain and cross-modal transfer from the audio-motor to the visual domain in musicians.

Additional evidence for cross-modal transfer from the motor to the visual domain stems from a body of research beyond the field of music. For instance, participants that had haptically explored unfamiliar objects while blindfolded were able to recognize these objects upon visual presentation (Bushnell and Baxt, 1999; Lacey and Campbell, 2006). To do so, haptic exploration must have generated a mental representation flexible enough to be transformed from one modality (e.g., the training or a universal modality) into another (i.e., the testing) modality. Further evidence comes from studies showing that a non-visual motor training procedure can improve visual perception (Casile and Giese, 2006; see also Hecht et al., 2001) and later visual presentation of trained movements results in stronger activity in a motor-related cortical network (Reithler et al., 2007).

Jointly, these studies not directly related to the music domain suggest that there is cross-modal transfer from the motor/haptic to the visual domain. Given that participants have viewed their own actions throughout their lives, the underlying sensory-motor contingencies have been trained over decades (see Brass et al., 2000, 2001; Iacoboni et al., 1999). The research on musicians also suggests cross-modal transfer for sensori-motor contingencies, such as playing an instrument. In these studies, however, the transfer occurs between modalities that have been engaged simultaneously during training phases. The research question we address here is whether such cross-modal interactions can develop rapidly, after only hours of practice for playing sequences on a piano keyboard, when modalities are *separated* during training sessions. Therefore, one modality is excluded during motor training sessions (i.e., either the auditory or visual modality, resulting in a purely visuo-motor and purely audio-motor training condition). Cross-modal transfer is later tested in that excluded domain. Here, we investigate by which brain structures a cross-modal transfer is mediated under such conditions. In the following, we outline two possible hypotheses:

(1) **Motor mediation hypothesis:** a number of authors have suggested that perception and action recruit common neural substrates (Hommel et al., 2001; Rizzolatti and Craighero, 2004), and that these substrates become shared between modalities due to Hebbian associations that are strengthened during training (Del Giudice et al., 2009; Keysers and Perrett, 2004). The activation of motor areas during action perception is consistent with the idea that action perception elicits covert sensori-motor processes that internally simulate the observed action (Rizzolatti and Craighero, 2004). Furthermore, studies on motor expertise show that activity in that network subserving action simulation increases with motor familiarity (Bangert et al., 2006; Lahav et al., 2007; Mutschler et al., 2007). Consequently, in case of transfer of information between the modalities, perception of trained compared to untrained sequences should elicit an association to a corresponding motor program, even if trained sequences were exercised in a different modality than the modality of actual perception. Such association could be reflected by stronger brain activity in *motor-related areas* in the contrast “perception of trained vs. untrained sequences”. However, for stronger brain activity to occur in that contrast, triggered single motor representations (i.e., one observed finger movement or one heard tone is related to the corresponding motor movement) must have been combined and linked to the motor program controlling the entire sequence of finger

movements. This may additionally activate sensory representations (of the training modality), but this is not necessary for recognition. Furthermore, this motor mediation hypothesis predicts higher recognition rates for visual than auditory presentation. Pre-existing visuo-motor associations (Brass et al., 2000, 2001; Iacoboni et al., 1999) may facilitate the activation of motor representations during observation or even already during audio-motor training sessions, and hence its (later) recognition. For auditory presentation, piano-naïve participants have no pre-existing associations between the sound of melodies and motor representations, and can thus only rely on the weaker associations established during audio-motor training sessions of other melodies. This might lead to weaker motor activations and recognition during listening to trained sequences learned in a visuo-motor condition.

(2) **Sensory mediation hypothesis:** the second possible hypothesis is based on the assumption that cross-modal transfer is facilitated via sensory areas, i.e., that connections between visual and auditory brain regions play a central role in recognition. Musicians show such sensory co-activation, e.g. in auditory areas when observing silent finger movements (Hasegawa et al., 2004; Haslinger et al., 2005). It is less clear, however, how such associations would develop when training sessions do not involve co-incident sound and vision of melodies. One possible scenario would be that the audio-motor training procedure would additionally elicit a visual image of finger movements or the visuo-motor training procedure would elicit auditory imagery. During perception of the trained vs. untrained movements, stronger activity in associated *sensory areas* should occur, since the trained movements would have a stronger resonance, e.g., through imagery processes during training procedures (Platel et al., 1997; Rossion et al., 2001; see for auditory imagery: Halpern and Zatorre, 1999; Zatorre and Halpern, 2005; for visual imagery: Roland and Gulyas, 1994). The sensory mediation hypothesis does not predict significant differences between conditions.

Methods

In order to study cross-modal transfer from the audio-motor to the visual domain, we trained non-musicians to play sequences on a piano via a purely audio-motor training procedure. In this audio-motor training condition, participants heard a melody and learned to reproduce it with the vision of their fingers occluded. After participants mastered the task, we showed them movies of the corresponding finger movements and compared these stimuli with the vision of untrained sequences in terms of brain activity (using fMRI scanning) and recognition (in a behavioral test after scanning). To study transfer from the visuo-motor to the auditory domain, sequences were trained in a purely visuo-motor fashion. In that training condition participants learned to imitate silent finger movements on a mute piano keyboard, but they were never hearing either the demonstrator's or their own piano sounds. We then played them the sound of either these trained or untrained melodies, and compared brain responses (using fMRI scanning) and recognition rates (in a behavioral test after scanning).

Participants

22 healthy volunteer non-musicians (mean age = 22 years; SD = 2.4 years; range 18–26 years; 11 female) participated in the current study. None of the participants had received any previous musical training (with the exception of music classes at school) or had played the piano before. All participants were right-handed according to the Edinburgh handedness inventory (Oldfield, 1971) and had normal vision or corrected to normal vision by contact lenses. They gave their written informed consent to participate, were naïve to the hypotheses, and received monetary compensation in return for participation. The experiment was performed in accordance with ethical standards

compliant with the declaration of Helsinki and had been approved by the Medical Ethical Commission of the University Medical Center Groningen, the Netherlands.

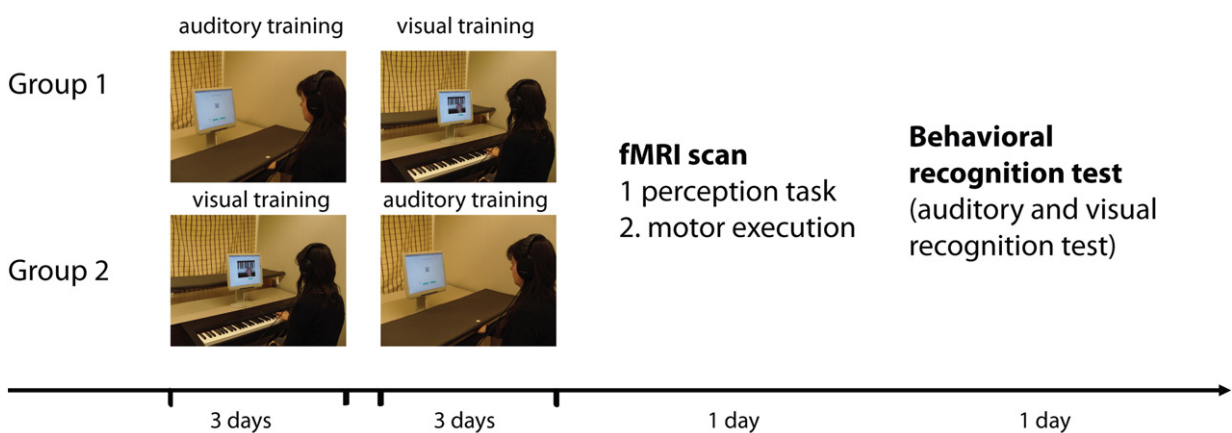
Overview of the study design

The entire procedure for each participant comprised two three-day motor training conditions in which they learned to perform different sequences on a piano keyboard; an fMRI scanning session and a behavioral recognition test after scanning (see Fig. 1A):

- (1) Motor training sessions I: first three days of motor training sessions (2 h each) in order to learn to perform three short key sequences (consisting of seven notes) within one training modality (audio-motor training condition for Group 1 (N = 11),

- containing half of the participants or visuo-motor training condition for Group 2 (N = 11), containing the other half).
- (2) Motor training sessions II: further three days (2 h each) of motor training sessions aimed at learning to perform three (new) key sequences within the other training modality (visuo-motor for Group 1 and audio-motor for Group 2).
- (3) fMRI scan: cross-modal transfer was tested by comparing brain activity during the perception of trained vs. untrained key sequences. Trained sequences were presented in the modality other than the training modality, i.e. observation of finger movements corresponding to melodies trained in the audio-motor condition and listening to melodies for which the corresponding finger sequence was trained in the visuo-motor condition. Since both participant groups were trained in both motor training conditions after six days, the fMRI data of all participants

A: Time course of training and testing



B: Time course of a trial of the perception task



C: Time course of a trial of the motor execution task

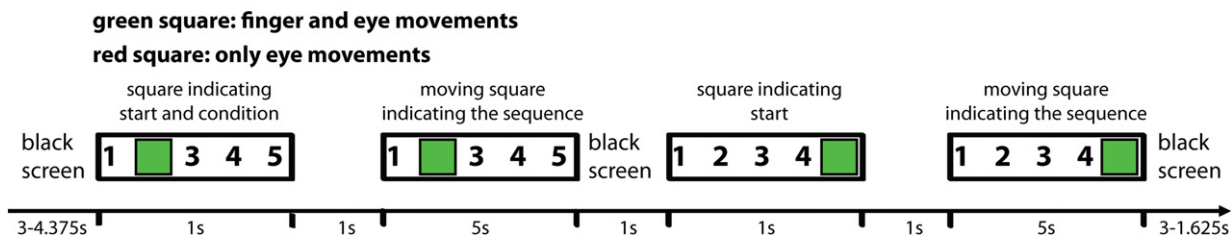


Fig. 1. Schematic representation of the time course of (A) training and testing procedure; (B) a trial of the perception task during scanning; and (C) of a trial of the motor execution task during scanning. (A) Each participant received a three-day audio-motor and three-day visuo-motor training procedure for learning to perform finger sequences on a piano keyboard. Half of the participants started with the audio-motor training procedure in which sequences were trained by blind listening to melodies and replaying. The other half of participants started with the visuo-motor training procedure in which sequences were trained by observation of silent finger movements and replaying. After six days of motor training sessions, an fMRI scanning session took place comprising a perception task (B) and a motor execution task (C). Data of all participants (N = 22) were analyzed as whole group regardless with which motor training condition they had started. After fMRI scanning, all participants performed a behavioral recognition test outside the scanner. (B) During the perception task in the fMRI scanner participants experienced either blocks in which they listened to two melodies or blocks in which they observed two finger movement sequences. In both kinds of blocks it was their task to pay attention whether the first or last sequence was played in a higher or lower position. (C) During the motor execution task in the fMRI scanner participants were required to perform either two sequences of finger (indicated by green squares) or eye movements (indicated by red squares). A stationary colored square first defined the experimental condition and start position of a sequence. Then the square started to move and covered numbers that corresponded to fingers or eye positions. Participants performed in each trial two movements according to these visual aids.

($N = 22$) were analyzed jointly as a whole group. Additionally, all participants performed a motor execution task.

- (4) Behavioral recognition test after scanning: all participants either observed or listened to trained and untrained sequences and had to judge whether they had learned to perform these sequences in one of the two motor training conditions.

Stimuli for the musical training procedure

Participants learned to perform three melodies using their right hand in a pure audio-motor training condition (e.g., Supplementary material Audio S1, Audio S2), and three melodies (i.e., key sequences) in a pure visuo-motor training condition (e.g., Supplementary material Video S1, Video S2). All sequences consisted of seven notes within three bars (four eighth notes, two quarter notes in the first two bars and one half note in the third bar; 2/4 measure; duration at target tempo, i.e., 120 beats per minute, bpm: 3000 ms; see Supplementary material Fig. S1). All piano key sequences were learned in two positions, using five adjacent white keys with each finger assigned to one piano key. One position ranged from C4–G4 and the other ranging from G4–D5 (finger assignments position1/position2: C4/G4 – thumb, D4/A4 – index finger, E4/B4 – middle finger, F4/C5 – ring finger, G5/D5 – little finger). This means that an identical motor pattern was necessary to perform both patterns but the corresponding sounds were lower or higher and the position of the hand for playing on the keyboard was shifted by five white keys (a musical interval of a fifth). Sequences had one of three different possible rhythms. The three visual, the three auditory and the three untrained sequences each contained one sequence of each rhythm, respectively (see Supplementary material Fig. S1). Each sequence utilized each of the five white keys on the piano keyboard once, and two keys were played twice. Half of the participants learned to play sequences 1–3 (Set 1, Supplementary material Fig. S1) in the visuo-motor training condition and the other three sequences (4–6, Set 2, Supplementary material Fig. S1) in the audio-motor training condition while the other half of the participants learned to play sequences 4–6 in the visuo-motor training condition and sequences 1–3 in the audio-motor training condition. Three further sequences (7–9, Set 3, Supplementary material Fig. S1) were used as untrained sequences presented during fMRI scanning.

Training conditions on the piano keyboard

Each of the non-musicians participated in two different motor training conditions in order to learn each three different sequences on the piano keyboard: (1) The visuo-motor training procedure required participants to learn by observing videos showing a real right hand in a bird's eye view performing silent finger movements on a piano keyboard (see Supplementary material Video S1, Video S2). Participants had to reproduce the sequences using a mute piano keyboard (Fig. 1A). (2) In the audio-motor training procedure participants learned by only listening to an audio track of a piano melody (hear Supplementary material Audio S1, Audio S2). They had to reproduce the sequence using the same piano keyboard as above, except that they could hear the notes they produced via headphones. Participants were prevented from seeing their own finger movements by a cover (see Fig. 1A). The audio-track of the demonstration was not accompanied by a video.

Each motor training condition took place on three successive days: on the first day participants learned to perform all three sequences in both positions in a slow tempo (60 bpm). On the second day, participants repeated the three sequences in the two positions in the slow tempo and learned to perform them in a fast tempo (120 bpm). On the third day participants repeated and consolidated the sequences in the fast tempo only to allow for high sensory-motor associations. Training procedure was standardized using a computer-based method with MIDI-based software (MaxMSP 5; see also Bangert et al., 2001) measuring the time, the number of

Table 1
Learning measures for piano training conditions.

	Audio-motor training condition	Visuo-motor training condition	Paired <i>t</i> -test (two-tailed)
Day 1 60 bpm			
Time (min)	60 ± 31	32 ± 28	$t(21) = 4.6; p < .001$
Presentation	148 ± 89	84 ± 75	$t(21) = 3.2; p < .01$
Trials	242 ± 150	121 ± 89	$t(21) = 4.4; p < .001$
Day 2 60 bpm			
Time (min)	18 ± 11	14 ± 11	$t(21) = 2.0; p = .05$
Presentation	40 ± 28	36 ± 29	$t(21) = 0.4; p = .66$
Trials	95 ± 67	68 ± 55	$t(21) = 2.5; p < .05$
Day 2 120 bpm			
Time (min)	17 ± 13	19 ± 15	$t(21) = -0.5; p = .60$
Presentation	50 ± 45	71 ± 61	$t(21) = -1.8; p = .09$
Trials	142 ± 113	132 ± 87	$t(21) = 0.4; p = .71$
Day 3 120 bpm			
Time (min)	6 ± 3	7 ± 6	$t(21) = 0.7; p = .51$
Presentation	19 ± 16	22 ± 24	$t(21) = -0.7; p = .49$
Trials	50 ± 28	56 ± 40	$t(21) = -0.6; p = .54$

Note: Values indicate mean ± standard deviation for the group of 22 participants comprising the sum of parameters for playing three sequences in two positions three times correctly, obtained from the computer training program applied the first time on a training day. Time (minutes needed), presentation (number of presentations) and trials (number of trials to reproduce the sequence) needed in order to reach the learning criterion of playing each sequences three times correctly. Results of paired (two-tailed) *t*-tests examining differences between the two training conditions are reported in the fourth column. bpm, beats per minute.

presentations of a model sequence and the number of trials a participant needed to reach the learning criterion (to play each sequence at least three times correctly). Please see the section “Methods: Training program and procedure” in the Supplementary material for a description of the computer-based training program and a detailed description of the training procedure on each training day. Each motor training condition comprised an over-learning procedure, which ensured that participants perceived and produced each sequence at least 50 times in each position at target tempo (120 bpm). Please note that Table 1 in the manuscript reports the number of presentations summed for the three sequences in two positions only for the first application of the computer program on a training day. At the end of each training session participants were expected to be able to play all sequences by heart (which was verified at the end of the second and the third training day).

Further behavioral measurements

Test for audio-motor associations

After the audio-motor training condition, we probed the degree to which participants had learned to associate arbitrary notes with a particular motor program. Participants heard a single tone and had to find the corresponding key on the keyboard (using the same computer program described in the Supplementary material). Time, number of presentations and number of trials needed to find the corresponding tone were registered. Participants had to reproduce 15 tones for each position. They started either with the 15 tones in the lower or higher position (balanced over participants and according to the start position in the first training session).

Recognition test

One day after the fMRI scanning, all participants underwent a behavioral recognition test. Participants first repeated the audio- and visuo-motor computer-based training program once (in the “ordered presentation mode”, see the section “Methods: Training program and procedure” in the Supplementary material, until they could play each sequence three times correctly) and were asked to repeat all sequences under free recall conditions. This was to ensure that the sequences of both training conditions were fresh in the participants' minds. The actual recognition test comprised an auditory and a visual recognition part with

the order of the two parts counterbalanced across participants and training groups.

Using an interactive computer presentation program (Presentation 11.3 software; Neurobehavioral Systems, <http://www.neurobs.com/>), participants could start the presentation of a sequence up to three times per mouse click in each trial. They were required to judge whether they have learned to play that sequence in one of the training sessions (“no” or “yes” – either in the visuo-motor or audio-motor training condition) and to report whether they are sure in their decision (2 = sure vs. 1 = not sure), resulting in a combined judgment with four possibilities: learned–sure; learned–not sure; not learned–sure; not learned–not sure. If participants judged that they had learned to play a sequence, they were asked to specify whether they learned to play it in the audio-motor or visuo-motor training condition. 18 Trials were performed in the visual (Vpres) and 18 trials in the auditory (Apres) presentation condition, corresponding to the three melodies learned in the audio-motor training condition (Atrain), the three melodies learned in the visuo-motor training condition (Vtrain) and three untrained melodies (Untrain), each of them presented once in the low and high position at the target tempo (120 bpm). Presentation order of different sequences and positions was randomized for each participant. Two of the untrained sequences were completely new for the participants and one untrained sequence was partially familiar, because it was also used as an untrained sequence during fMRI scanning session.

Debriefing

On motor training days and on the scanning day, a debriefing session took place. Participants were asked which strategy they adopted to learn to play the piano melodies and how they solved the task during scanning. Additionally, ratings for concentration, motivation, self-performance, etc. were acquired on a four-point rating scale and two questionnaires were applied. One questionnaire determined preferred learning strategies (VARK, Fleming and Mills, 1992) and the other one assessed aspects of empathy (Interpersonal Reactivity Index, Davis, 1983).

Experimental procedure during fMRI scanning

During the fMRI experiment, participants laid in a supine position on the scanner bed, with the right hand resting on the scanner compatible response box. Written instructions were projected by an LCD projector onto a screen placed behind the participant's head. The screen was viewed via a mirror on the top of the head coil. All auditory stimuli were presented via scanner-compatible headphones (MRconfon GmbH, Magdeburg, Germany; <http://www.mr-confon.de/>). Stimulus delivery was controlled by Presentation 11.3 software running on a computer.

fMRI scanning took place after audio- and visuo-motor training sessions and was divided into four runs for the perception task and one run for the motor task. There was an auditory and a visual perception part within each of the first four runs. Whether a run started with the visual or the auditory part was counterbalanced between and within participants. Participants observed within each trial in the visual part two sequences successively in two different positions, and listened to two melodies presented in two different positions in each trial of the auditory part (see Fig. 1B). The motor pattern required to perform these sequences could be either the same or different. To ensure that participants paid a comparable amount of attention to the stimuli, it was their task to identify whether the first or the last sequence was played in a higher or lower position in all trials (catch task). In a third of all trials they were requested to indicate their response by button press (the visual signal to do so appeared randomly during the run). Participants were instructed always to attend to all sequences in their entirety (as it was not possible to anticipate whether or not they needed to produce a response). Due to the possible

need to respond, participants' right hands were held in a fixed position on a response box during scanning. Additionally they were instructed not to move their fingers or any other part of their bodies during scanning. In subsequent debriefing, most participants confirmed that they did not move their fingers during scanning (20 out of 22 participants indicated they had not moved their fingers, one participant said he had sometimes pressed the buttons accidentally and another participant claimed he had sometimes stretched his fingers).

The melodies presented during scanning in the perception runs were either trained or untrained (see Supplementary material Fig. S1). This resulted in a 3×2 factorial design: training condition (audio-motor/visuo-motor/untrained), presentation modality (auditory/visual) for the whole group (22 participants). For each of these conditions there were 24 trials, presented in a randomized order over different runs. Each trial (see Fig. 1B) had a length of 15 s, starting with a fixation cross (variable length, 1000–2375 ms). After a further 500 ms black screen, two sequences were presented (120 bpm, each video /audio file was 4000 ms long, i.e., 3000 ms for the sequence and a pause of 500 ms at the beginning and the end of each sequence), followed by a variable inter trial interval of 4125–5500 ms in which either a black screen was presented or in a third of the trials a question appeared for 3000 ms (position task, four possible questions: “last/first pattern in a higher/lower position?”). Participants answered with a press of the right hand index finger for lower position or the right hand middle finger for higher position on the response box.

Additionally, after the four perception runs, a fifth run took place for localizing regions involved in the motor control of the fingers. Participants had to perform eye movements and finger (+ eye) movements matching the movements of the trained key sequences (see Fig. 1C). A detailed description of the task can be found in the section “Methods: Motor execution task” in the Supplementary material. Data from the finger (+eye) movement condition were contrasted against those from the eye movement condition, which permitted us to isolate brain regions selectively involved in the control of sequences of finger movements after removing effects of unspecific executive and visual processing that occurs in common with the eye-movement condition. This activation map was used to identify brain regions that are active during the perception and execution of finger movements.

MRI acquisition and data analysis

MR scans were performed on a 3 Tesla Philips Intera MR-scanner with an eight-channel synergy SENSE head coil. In the scanning session, four runs of 370 and one run of 340 functional whole brain images, sensitive to the blood oxygenation level dependent (BOLD) signal, were acquired using a T2*-weighted echo planar imaging sequence. Each volume consisted of 29 AC-PC aligned slices with the following parameters: voxel size 3.5 mm×3.5 mm and a slice thickness of 4.0 mm without interslice gap; repetition time (TR) 1500 ms, echo time (TE) 28 ms, flip angle 70°, field of view (FOV) 224 mm×224 mm×116 mm. Interleaved ascending image acquisition was used in which odd numbered slices were acquired before even numbered slices. During scanning, parallel imaging and a soft-tone option (which decreases the scanner noise) was used and participants followed instructions to stay as still as possible. Additionally, a set of high-resolution T1 weighted anatomical images was acquired on a different scanning day (voxel size 1 mm×1 mm×1 mm, flip angle 8°, FOV 256 mm×224 mm×160 mm, 160 slices).

fMRI data were analyzed using SPM8 (<http://www.fil.ion.ucl.ac.uk/spm/>) implemented in Matlab R14 (The Mathworks INC., <http://www.mathworks.com/>). All functional images were pre-processed by applying slice time correction and realigning all volumes of each subject to the first functional volume. Functional images were co-registered to the 3D anatomical image of the participant. The 3D anatomical image was segmented and the gray matter segment was normalized to a gray matter template corresponding to the Montreal Neurological Institute

(MNI) brain template and obtained parameters were used for normalization of the functional data. The voxel dimensions of each reconstructed functional scan were $3\text{ mm} \times 3\text{ mm} \times 3\text{ mm}$. Finally, functional images were spatially smoothed with an 8 mm full-width half-maximum Gaussian filter.

In the first level analysis, pre-processed images of each participant were analyzed with a General Linear Model. For the motor execution run, two predictors of interest were modeled in the first level analysis: eye movements and finger (+eye) movements, each as a boxcar function with a length of 14 s convolved with the canonical hemodynamic response function and the temporal derivative. A high pass filter of 128 s was used to remove low frequency noise. A contrast “finger vs. eye movements” was calculated on the first level and the contrast images of each participant were submitted to a second level random effects analysis using a one-sample *t*-test. For the four perception runs, the GLM model of the first level comprised 6 predictors of interest covering (1) visually presented sequences learned in the visuo-motor training condition (VpresVtrain); (2) visually presented sequences learned in the audio-motor training condition (VpresAtrain); (3) visually presented sequences, untrained (VpresUntrain); (4) auditorily presented sequences learned in the visuo-motor training condition (ApresVtrain); (5) auditorily presented sequences learned in the audio-motor training condition (ApresAtrain); (6) auditorily presented sequences, untrained (ApresUntrain). Furthermore the response for the attention task was modeled as a predictor of no interest. All predictors of interest were modeled as boxcar functions with a length of 8 s convolved with the canonical hemodynamic response function and the temporal derivative. Low frequency drifts from the perceptual functional runs were removed using a high pass filter of 516 s and a correction for autocorrelation (AR(1)) was applied. In the first level analyses, simple contrasts were generated for each predictor of interest vs. un-modeled resting baseline for each individual participant. The results of the contrasts of all participants ($N=22$) were entered into a one-way analysis of variance (ANOVA) model for a second-level analysis. In the second level, firstly, two contrasts of main interest were calculated to compare sequences that differed in motor familiarity but not perceptual familiarity: (1) visual presentation of finger movements corresponding to melodies learned in the audio-motor training condition vs. visual presentation of untrained finger movements (VpresAtrain vs. VpresUntrain); (2) auditory presentation of melodies corresponding to sequences learned in the visuo-motor training condition vs. auditory presentation of untrained melodies (ApresVtrain vs. ApresUntrain). Secondly, two additional contrasts comparing sequences that differed in the need for cross-modal transfer and perceptual familiarity but not in their motor familiarity were calculated: (3) visual presentation of finger movements corresponding to melodies learned in the audio-motor training condition vs. visual presentation of finger movements learned in the visuo-motor training condition (VpresAtrain vs. VpresVtrain); (4) auditory presentation of melodies corresponding to sequences learned in the visuo-motor training condition vs. auditory presentation of melodies learned in the audio-motor training condition (ApresVtrain vs. ApresAtrain). All these contrasts were first masked inclusively with the motor execution map (thresholded at $p < 0.001$, uncorrected) and for further exploration calculated without mask. For identified areas, we report activations that were significant at $p < 0.05$, corrected for multiple comparisons (using the False Discovery Rate, FDR). Additionally, we report the results at the significance level of $p < 0.001$, uncorrected for multiple comparisons. Activation cluster with sizes $k > 5$ voxel are considered.

Displays of activations were created by means of the software package MRICron (<http://www.sph.sc.edu/comd/rorden/mricron/>) by superimposing SPM *t*-maps resulting from the second level analysis maps on a mean anatomy, obtained by averaging the normalized anatomies of all 22 participants. Labeling of activation clusters was done with the anatomy toolbox (http://www.fz-juelich.de/inm/inm-1/spm_anatomy_toolbox/) and brain atlases.

Results

Behavioral data

Learning measures of piano training conditions

Learning times, number of presentations and number of trials needed to learn and perform all three melodies in the two positions three times correctly are reported in Table 1 for the audio-motor and the visuo-motor training condition for all participants ($N=22$; Supplementary Material Fig. S2 illustrates the learning times). Please note that these values refer to learning indices obtained from the first training part of a day (repetitions of the training program served for consolidation, i.e., over-learning effects, and are not reported in detail). *t*-Tests comparing the indices for the audio-motor and visuo-motor training condition revealed significant differences for the first training day (learning the sequences in a slow tempo) for all learning indices (see Table 1) indicating faster learning in the visuo-motor training condition while needing less presentations and trials. On the second training day, participants needed more trials in the audio-motor training condition than in the visuo-motor training condition for repeating the sequences in the slow tempo, but no significant differences were found between the training conditions for learning time or number of presentations. No significant differences between training conditions were found in any learning indices for learning the fast tempo on the second training day. Finally and most importantly, no differences on the third training day were found for the audio-motor and visuo-motor training condition. Please note that the unequal number of presentations for sequences at target tempo in the audio-motor and visuo-motor training condition was balanced with an over-learning procedure.

A differential analysis of learning measures according to the training order, whether participants started with the audio-motor or visuo-motor training condition can be found in the section “Results: Behavioral data – learning times” and Table S1 in the Supplementary material. Furthermore, results concerning individual learning preferences can be found in the section “Results: Behavioral data – learning preferences” in the Supplementary material.

Test for audio-motor associations

During the auditory tone test after the audio-motor training sessions, participants needed on average \pm SD: 1.2 ± 0.2 presentations of the tone, 1.7 ± 0.4 trials and 7.9 ± 1.9 s time for finding the correct key.

Recognition test

Table 2 summarizes the results of the recognition test. Fig. 2A illustrates the recognition rates for visual and Fig. 2B for auditory presentations. A two (presentation mode: auditory/visual) \times three (training condition: same training modality as presentation = no transfer, different training modality as presentation = cross-modal transfer, no training sessions) repeated measures ANOVA on the percentage judged as trained revealed significant main effects of presentation mode ($F(1,21) = 8.0$, $p < .05$), training condition ($F(2,42) = 176.5$, $p < .001$) and a significant interaction ($F(2,42) = 28.4$, $p < .001$).

Two-sample *t*-tests for recognizing sequences within the modality of the training procedure (no cross-modal transfer is necessary) revealed that participants ability to recognize auditory presentations of audio-motor trained sequences did not differ from their ability to recognize visual presentations of visuo-motor trained sequences (ApresAtrain = VpresVtrain: $t(21) = 1.4$; $p = .16$). However, in the recognition conditions where cross-modal transfer was necessary, participants were better in recognizing visual presentations of audio-motor trained sequences than recognizing auditory presentations of visuo-motor trained sequences (VpresAtrain > ApresVtrain: $t(21) = 2.3$; $p < .05$). Untrained sequences were more often falsely judged to be trained by auditory than visual presentation (ApresUntrain > VpresUntrain: $t(21) = 5.8$; $p < .001$).

In both presentation modalities, participants were better at recognizing melodies presented within their original training modality than within the other modality (VpresVtrain > VpresAtrain: $t(21) = 2.5$, $p < .05$, Fig. 2A, and ApresAtrain > ApresVtrain: $t(21) = 4.8$; $p < .001$, Fig. 2B). However this difference was larger for the auditory (ApresAtrain – ApresVtrain: $22.7 \pm 22.2\%$) than visual (VpresVtrain – VpresAtrain: $9.9 \pm 18.3\%$) presentation (t -test of the difference contrast, i.e., the interaction ([ApresAtrain – ApresVtrain] – [VpresVtrain – VpresAtrain]): $t(21) = 2.5$, $p < .05$). Importantly, listening to visuo-motor trained melodies led to more frequent labeling as ‘trained’ than listening to untrained melodies although both were perceptually unfamiliar for participants in the auditory modality (ApresVtrain > ApresUntrain: $t(21) = 5.2$, $p < .001$; Fig. 2B). Viewed sequences from melodies of the audio-motor training condition were also labeled as ‘trained’ more often than viewed sequences from melodies of the untrained condition (VpresAtrain > VpresUntrain: $t(21) = 15.3$, $p < .001$, Fig. 2A). But this difference (VpresAtrain – VpresUntrain: $81.1 \pm 24.8\%$) was larger than that of the corresponding difference in the auditory (ApresVtrain – ApresUntrain: $37.1 \pm 33.7\%$) presentation testing (t -test of the difference contrast, i.e., the interaction ([VpresAtrain – VpresUntrain] – [ApresVtrain – ApresUntrain]): $t = 6.3$, $p < .001$).

Please find in the section “Results: Behavioral data – recognition test” (a) the values and statistical analyses of the recognition rates of untrained sequences that are corrected by excluding the melody that was used during fMRI scanning (see also Supplementary material Table S2) and a detailed analyses of (b) reaction times, (c) confidence ratings and (d) the judgment about the training modality. Here we summarize briefly the results of the recognition test for the conditions in which cross-modal transfer was necessary (see Table 2): participants were faster ($t(21) = 4.4$; $p < .001$) and more confident ($t(21) = -5.3$; $p < .001$) in their judgment, whether they learned to play a melody and more often correct in the judgment about the modality of the training condition ($t(21) = -2.6$; $p < .05$) when observing finger movements, learned in the audio-motor training condition (VpresAtrain) than when listening to melodies learned in the visuo-motor training condition (ApresVtrain).

Table 2
Behavioral recognition test.

	APres	VPres	Paired t -test
No-transfer	Atrain	Vtrain	
% judged as trained	100.0 ± 0.0	98.5 ± 4.9	$t(21) = 1.4$; $p = .16$
Reaction time (ms)	1313 ± 369	1414 ± 268	$t(21) = -1.3$; $p = .22$
Confidence	1.94 ± 0.10	1.94 ± 0.14	$t(21) = 0$;
% correct training modality	97.0 ± 8.4	90.9 ± 16.0	$t(21) = 1.7$; $p = .10$
Cross-modal transfer	Vtrain	Atrain	
% judged as trained	77.3 ± 22.1	88.6 ± 16.6	$t(21) = -2.3$; $p < .05$
Reaction time (ms)	2943 ± 1855	1880 ± 1380	$t(21) = 4.4$; $p < .001$
Confidence	1.46 ± 0.30	1.83 ± 0.27	$t(21) = -5.3$; $p < .001$
% correct training modality	83.4 ± 23.7	97.3 ± 7.0	$t(21) = -2.6$; $p < .05$
Untrain			
% judged as trained	40.2 ± 27.1	7.6 ± 14.3	$t(21) = 5.8$; $p < .001$
Reaction time (ms)	3623 ± 3270	2348 ± 1643	$t(21) = 1.6$; $p = .12$
Confidence	1.39 ± 0.31	1.74 ± 0.28	$t(21) = -4.0$; $p < .01$

Note: Values indicate mean ± standard deviation. Each cell represents the results of only six sequences (three melodies presented in two positions each). Second and third columns – testing modality (Apres: auditory presentation, Vpres: visual presentation). Fourth column: t -values, degrees of freedom and p -values are reported for paired two-tailed t -test, comparing auditory and visual presentation. Rows: “No-transfer” – melodies trained in the presentation modality; Cross-modal transfer – melodies trained in a different modality (Atrain: audio-motor trained, Vtrain: visuo-motor trained); Untrain – presentation of untrained sequences. % judged trained: percentage of 3×2 sequences that were judged to be learned; Reaction time: reaction time in milliseconds for the judgment trained or untrained (which was combined with the confidence rating); Confidence: confidence of judgment (2 = sure, 1 = not sure); % correct training modality: if a melody was considered as being trained, participants had to specify in which modality they had trained that melody – number correctly assigned to modality/number of correctly assigned as trained * 100.

Debriefing

At the end of the last session of each motor training condition we debriefed the participants for strategies concerning imagery during training sessions. We did so only at the end of the motor training sessions, in order not to induce or propose strategies for motor learning during an ongoing training session. The majority of participants (16 out of 22 participants) answered to have not imagined finger movements during the audio-motor training condition. Only three out of 22 participants reported to have imagined finger movements during the audio-motor training condition, one participant did imagine finger movements sometimes and two participants didn't answer the question. Thus, recognition rates during observation of finger movements are unlikely to be explained by imagery of finger movements during the audio-motor training sessions. The same debriefing after the visuo-motor training condition revealed that the majority of participants (18 out of 22 participants) imagined a sound during the visuo-motor training condition. More specifically this imagery involved tones for 10 participants (which might not be necessarily related to the corresponding melody; 8 of these 10 participants had completed the audio-motor training sessions before) or a rhythm imagery (like sub-vocal “bam bam”) for 8 participants (3 of these participants had completed the audio-motor training sessions before). Four participants answered not to have had any sound imagery during the visuo-motor training procedure.

Task performance during scanning sessions (catch task)

Analysis of participants' responses during scanning revealed that they could recognize whether a sequence was in the higher or lower position with a high correct response rate (mean ± SD) during the visual presentation condition ($99.4 \pm 1.5\%$) and the auditory presentation condition ($91.9 \pm 9.5\%$), respectively. These correct response rates differed significantly ($t(21) = 4.0$; $p < .01$). Thus, the visual catch task was easier than the auditory discrimination for the participants. However, the correct response rates were still relatively high in the auditory task, indicating that participants paid attention to the sequences during the perception. It should be noted that the catch task regarding the position of sequences is unrelated to aims of the present study to investigate cross-modal transfer of trained sequences, since the catch task does not require (conscious) recognition of trained sequences. Thus, the analyses of brain activations during perception of trained vs. untrained sequences (for cross-modal transfer) identify rather automatic processing.

fMRI data

Motor execution task

Analysis of participants' hemodynamic responses during “execution of finger movements compared to execution of eye movements” revealed stronger activity in a network of brain areas comprising the left primary motor and bilateral somatosensory area, left premotor area, left insula, bilateral Brodmann Area (BA) 44, left rolandic operculum, left auditory cortex, bilateral inferior and superior parietal lobe, bilateral middle frontal gyrus, right fusiform gyrus and the cerebellum (see Table 3 and Supplementary material Fig. S3). These brain activations occurred at a significance level of $p < .001$, uncorrected ($t(21) = 3.5$) and all activations were also significant at $p < .05$, FDR corrected ($t(21) = 2.4$; please note that FDR correction in that contrast entails a more liberal significance level). This activation map served as a mask for identifying areas that were also activated during the perception task, thus ensuring that identified areas are also active during the execution of finger movements.

Observation of finger movements

The contrast for the cross-modal transfer, “observation of finger movements corresponding to audio-motor trained melodies vs. observation of untrained finger movement sequences (VpresAtrain vs. VpresUntrain)” calculated for the whole group of participants

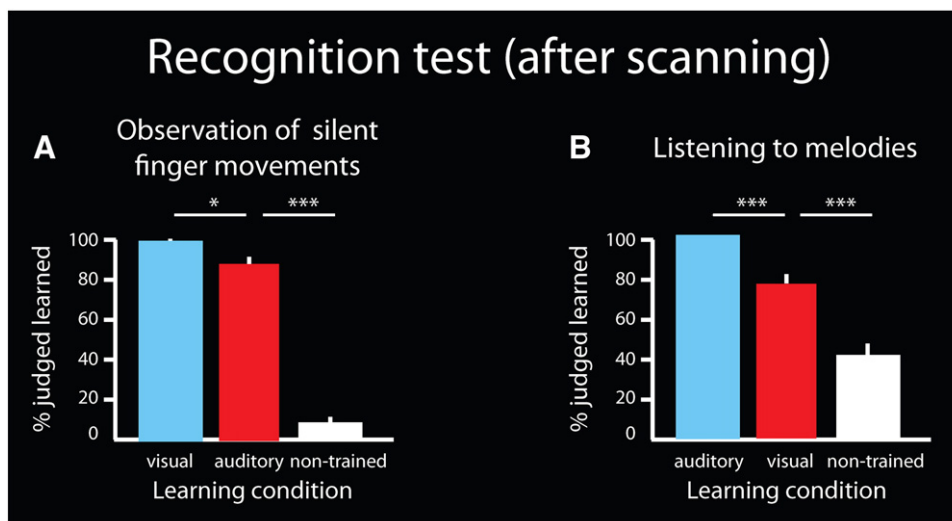


Fig. 2. Behavioral recognition test performed after the fMRI scanning session. (A) Observation of each three finger movements (in two positions) trained in the visuo-motor training condition (blue, VpresVtrain, no cross-modal transfer necessary for recognition), in the audio-motor training condition (red, VpresAtrain, cross-modal transfer necessary for recognition) and untrained sequences (white, VpresUntrain) showing the percentage of responses across participants reporting to have learned a sequence. (B) Listening to melodies trained in the audio-motor training condition (blue, ApresAtrain, no cross-modal transfer necessary for recognition), in the visuo-motor training condition (red, ApresVtrain, cross-modal transfer necessary for recognition) and untrained sequences (white, ApresUntrain) showing the percentage of responses across participants reporting to have learned a sequence. Error bars in (A) and (B) represent the standard error of mean. Asterisks indicate the significance level between conditions: * for $p < .05$; *** for $p < .001$.

($N = 22$) revealed a significant activation cluster in the left rolandic operculum at a significance level of $p < .05$, FDR corrected ($t(126) = 3.7$, Table 4, Fig. 3). This contrast was masked with the activation map for execution of finger movements. Thus this identified area was also active during the execution of finger movements. When lowering the threshold to $p < .001$ uncorrected ($t(126) = 3.2$) in that masked contrast, this analysis revealed that the activation cluster of the rolandic operculum extended into BA44. Further activation clusters were identified in the left dorsal premotor areas, the right cerebellum and left putamen at this lowered threshold (Table 4).

In order to test whether areas outside of those active during the execution of finger movements are involved in the cross-modal

transfer from the audio-motor training condition to the visual presentation domain, we calculated the same contrast, VpresAtrain vs. VpresUntrain, in a whole brain analysis, $p < .001$, uncorrected ($t(126) = 3.2$) for all participants ($N = 22$). This analysis revealed additionally to the above mentioned areas activity in the bilateral anterior cingulate cortex, the left superior, middle and superior medial frontal gyri, the right insula, bilateral hippocampus, right caudate and the right superior parietal lobe (see Fig. 4, Table 5).

The contrast “observation of finger movements corresponding to audio-motor trained melodies vs. observation of finger movement sequences learned in the visuo-motor training condition (VpresAtrain vs. VpresVtrain)” was firstly calculated with a significance level of $p < .05$, FDR corrected ($t(126) = 2.6$, Table 6) for the whole group of participants ($N = 22$) and masked with the motor execution map. This analysis revealed an activation cluster in the bilateral BA44 extending into the rolandic operculum in the left hemisphere, the left anterior insula, the bilateral dorsal premotor area, the supplementary motor area, the middle cingulate cortex, the bilateral inferior parietal lobe and the bilateral somatosensory cortex, the right superior parietal lobe and the brainstem. Adopting a more conservative significance level of $p < .001$, uncorrected, in that contrast ($t(126) = 3.2$) resulted in a reduction in cluster sizes in all clusters (reported at the level $p < .05$, FDR corrected). As a consequence of this, the brain activation in the cluster in the left BA44 did not reach into the left rolandic operculum and the cluster of the left anterior insula in addition to one of the two clusters in the inferior parietal lobe did not survive that significance criterion. The contrast VpresAtrain vs. VpresVtrain in a whole brain analysis (without masking with the motor execution map) at $p < .001$, uncorrected ($t(126) = 3.2$) revealed two additional frontal activation clusters but no activation in auditory sensory areas (Supplementary material Table S3).

Listening to melodies

The contrasts a) for the cross-modal transfer, “listening to melodies corresponding to visuo-motor trained sequences vs. listening to untrained melodies (ApresVtrain vs. ApresUntrain)” and b) “listening to melodies corresponding to visuo-motor trained sequences vs. listening to melodies learned in the audio-motor training condition

Table 3
Brain activations during “execution of finger vs. eye movements”.

Anatomical region	Hemi-sphere	MNI coordinates			Cluster size	Z-score
		x	y	z		
Cluster I					7435	
Primary motor area	L	-36	-31	49		7.10
Somatosensory Area	L	-39	-28	49		6.88
IPL	L	-39	-34	34		6.22
Superior parietal lobe	L	-24	-49	61		5.85
Auditory cortex	L	-45	-25	13		5.32
Premotor area	L	-27	-13	61		7.37
BA44	L	-60	8	7		4.45
Anterior insula	L	-45	2	4		4.80
Rolandic operculum	L	-54	5	4		5.53
Cerebellum	R	21	-55	-23		7.56
Cluster II					786	
IPL/intraparietal sulcus	R	39	-37	43		5.65
Somatosensory area (BA2)	R	51	-25	43		5.19
Somatosensory area (BA1)	R	57	-16	40		4.87
BA44	R	57	14	34	152	4.39
Superior parietal lobe, precuneus	R	12	-67	46	63	4.01
Middle frontal gyrus	L	-30	32	31	48	4.07
Fusiform gyrus	R	48	-67	-17	11	3.46
Middle frontal gyrus	R	39	50	22	7	3.43

Note: The values shown are Montreal Neurological Institute (MNI) coordinates for significant activation maxima of clusters in the random effects analyses ($p < .001$, uncorrected). IPL, inferior parietal lobe; BA, Brodmann area; R, right; L, left.

Table 4

Brain activations during observation of finger movements corresponding to melodies learned in the audio-motor training condition vs. untrained sequences (VpresAtrain vs. VpresUntrain, masked with motor execution map).

Anatomical region	Hemi-sphere	MNI coordinates			Cluster size	Z-score
		x	y	z		
<i>VpresAtrain vs. VpresUntrain (p < .05, FDR corrected)</i>						
Rolandic operculum	L	-45	-1	7	22	4.32
<i>VpresAtrain vs. VpresUntrain (p < .001, uncorrected)</i>						
Rolandic operculum/BA44	L	-45	-1	7	72	4.32
BA44	L	-48	11	4		3.34
Cerebellum	R	24	-55	-23	20	3.56
Putamen	L	-30	-19	-5	11	3.46
Dorsal premotor area	L	-18	2	61	7	3.36
Extra-nuclear, putamen	L	-21	5	-11	7	3.73

Note: The values shown are Montreal Neurological Institute (MNI) coordinates for significant activation maxima of clusters. R, right; L, left, BA, Brodmann area, VpresAtrain, visual presentation of audio-motor trained sequences; VpresUntrain, visual presentation of untrained sequences.

(ApresVtrain vs. ApresAtrain)", calculated for all participants (N = 22), revealed no significant activations, neither at the level $p < .05$, FDR corrected nor at $p < .001$ uncorrected. This held true for limiting the contrasts to the areas that were also active during the execution of finger movements, as well as when calculating this contrast in a whole brain analysis without masking.

Discussion

Here we investigated the cross-modal transfer of movement patterns necessary to perform melodies on a piano keyboard. Specifically, we examined whether or not sequences learned without the visual modality would be recognized in the visual modality, and whether or not sequences learned without the auditory modality would be recognized in the auditory modality. Initial learning was faster in the visuo-motor training condition than in the audio-motor training condition. The entire motor training procedure, together with an over-learning phase, aimed to compensate for these initial learning differences by ensuring comparable motor and perceptual familiarity for audio-motor and visuo-motor trained sequences. However, participants recognized the vision of corresponding finger movements of audio-motor trained melodies more accurately, more rapidly, and more confidently than the corresponding sound of visuo-motor trained sequences. That information transferred more from the audio-motor training modality to

the visual testing modality than in the direction from the visuo-motor training modality to auditory testing modality, was mirrored by the pattern of brain activity. Stronger activation in the left rolandic operculum was found reliably for observing finger movements corresponding to audio-motor trained melodies compared to untrained ones (VpresAtrain vs. VpresUntrain), suggesting that this part of the neural representations generated during the audio-motor training sessions had become available to the visual modality. The same region, extending into BA44, was also found to be more active for observing finger movements corresponding to audio-motor trained melodies compared to those learned in the visuo-motor training condition (VpresAtrain vs. VpresVtrain) suggesting that the region may play an active role in cross-modal transfer. In contrast, no significant differences in brain activation were measured between listening to the corresponding sound of visuo-motor trained compared to untrained sequences, suggesting that neural representations generated in the visuo-motor training sessions did not become available enough to the auditory modality to generate significant differential brain activation. In the following we will discuss the pattern of results we observed (a) during training sessions, (b) in the recognition test and (c) in brain activations.

Analysis of the performance parameters during the audio-motor and the visuo-motor training conditions revealed that participants needed fewer trials, fewer presentations and less time in the visuo-motor than in the audio-motor training condition to reach the learning criterion on the first training day, during which participants learned to play sequences at the slow tempo. These results show that acquiring the appropriate motor sequences is easier in the visuo-motor than in the audio-motor training condition. An inherent difference between the training conditions is that, in addition to learning new sequences of movements, the audio-motor but not the visuo-motor condition also requires acquiring the mapping between specific notes and specific finger movements. In the visuo-motor training condition, pre-existing associations between seeing a specific finger move and moving the corresponding finger exist (Brass et al., 2000, 2001; Iacoboni et al., 1999), because all our participants have seen their own fingers "in action", for example, when pressing keys while typing on a computer. In contrast, our piano-naïve participants lack associations between the sound of particular notes and the finger movements to produce them. The test for audio-motor associations after the audio-motor training condition confirmed that participants learned these note-finger associations during the course of the training procedure, although these mappings did not yet lead to perfect performance (often more than one trial was needed to find the correct key for a perceived tone). The need to establish these note-finger associations must have slowed the acquisition of the motor sequences in the audio-motor compared to the visuo-motor training condition. For the second training day, differences in learning parameters between training conditions were only present in the number of trials needed to play each sequence three times correctly in the slow tempo. For learning the fast tempo, no differences in the performance parameters appeared between the two training conditions. Participants' subjective reports revealed that they evaluated learning the rhythm at the fast tempo to be more difficult in the visuo-motor than in the audio-motor training condition. This is consistent with the claim that audition provides better and more accurate temporal resolution (Kato and Konishi, 2006; Repp and Penel, 2002, 2004). Importantly, no differences in performance parameters were found on the third training day, when participants repeated the trained sequences at the target tempo. Thus the sequence of finger movements must have been already well established in the motor programs of participants in both training conditions. Furthermore, participants were able to play all sequences by heart at the end of each training condition. This suggests that the representation of the motor program was well consolidated and became part of participants' motor repertoire, which could be recalled without an additional perceptual

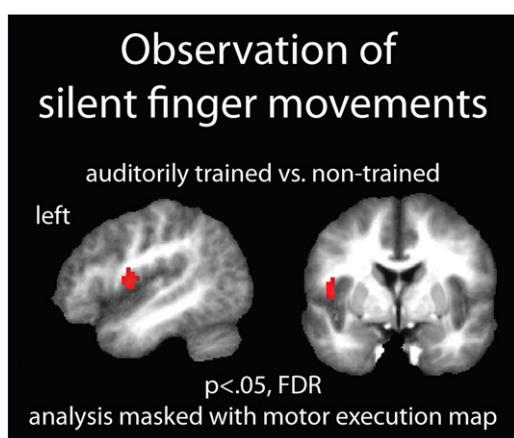


Fig. 3. Observation of finger movements. BOLD activation patterns based on the contrast of interest "observation of finger movements corresponding to audio-motor trained vs. untrained melodies (VpresAtrain vs. VpresUntrain)" masked with the motor execution map at significance level $p < .05$, FDR corrected. Statistical parametric map (SPM) is superimposed on a mean brain based on the normalized anatomies of the 22 participants.

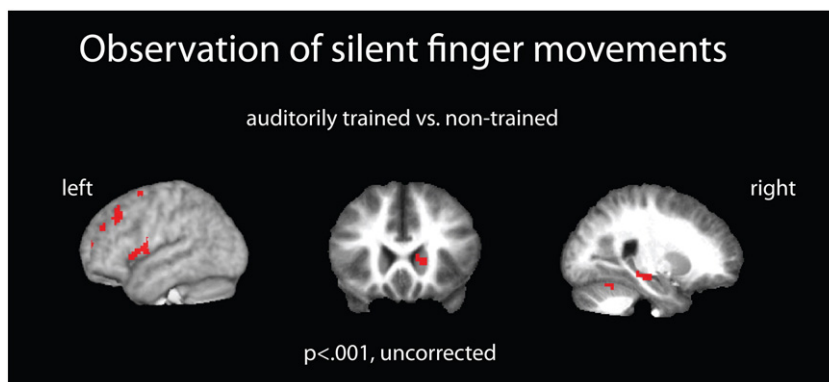


Fig. 4. Observation of finger movements. BOLD activation patterns based on the contrast of interest “observation of finger movements corresponding to audio-motor trained vs. untrained melodies (VpresAtrain vs. VpresUntrain)” for a whole brain analysis (unmasked) at significance level $p < .001$, uncorrected. Statistical parametric map (SPM) is superimposed on a mean brain based on the normalized anatomies of the 22 participants.

external cue presenting the sequence. Consequently, training sequences in the audio-motor and visuo-motor condition had led to similar levels of motor performance, suggesting comparability in motor familiarity for auditorily and visually trained sequences. The over-learning procedure further ensured similar perceptual levels of familiarity with all trained melodies (participants perceived at least 50 presentations at target tempo per sequence and per position). Hence, all trained sequences were highly familiar, both motorically and perceptually within the modality of training condition, by the end of training sessions. This is important for interpreting differences in cross-modal transfer.

The behavioral recognition test (after the training procedure and scanning) corroborates the conclusion of comparability in familiarity within the training condition: there were no significant differences in terms of recognizing sequences as having been trained or not between auditory and visual presentation of sequences when presented in the training modality. The recognition test however showed that the two training schemes lead to different amounts of cross-modal transfer. When participants had to recognize sequences presented in the modality that had been excluded from training procedure, recognition was faster, more accurate, and more confident when observing finger movements corresponding to melodies acquired blindly in the audio-motor training condition than when listening to

melodies corresponding to sequences acquired in the silent visuo-motor training condition. This implies that representations created in the audio-motor training condition were more accessible to vision than representations created in the visuo-motor training condition to audition. Thus, the visual percept of a series of finger movements could be related to the sequences trained and consolidated in the blind audio-motor training condition. One might question whether such a visual percept was already facilitated by imagery of finger movements during audio-motor training sessions. However, the debriefing of participants on possible imagery strategies during the audio-motor training condition did not support that assumption. Rather, an immediate matching of the motor sequence with a visual percept in the moment of observation seems to be more plausible. Such matching might have benefited from the above-discussed pre-existing associations between perceiving finger motion and the corresponding motor program (Brass et al., 2000, 2001; Iacoboni et al., 1999). Although the majority of participants reported the use of auditory images during the visuo-motor training condition, the matching of a heard melody to a corresponding motor sequence was less reliable during listening. However, this appears less contradictory when considering participants' reports that their auditory images did not necessarily match with the true melody (since non-musicians had no preexisting tone–finger or finger–tone associations) and that these auditory images rather represented rhythmic elements. Furthermore, the test for audio-motor associations

Table 5
Brain activations during observation of finger movements corresponding to melodies learned in the audio-motor training condition vs. untrained sequences (VpresAtrain vs. VpresUntrain, $p < .001$, uncorrected, unmasked).

Anatomical region	Hemi-sphere	MNI coordinates			Cluster size	Z-score
		x	y	z		
Rolandic operculum/BA44	L	-45	-1	7	117	4.34
Anterior cingulate cortex	L	-9	38	22	48	4.24
Putamen (partly covering)	L	-21	8	-11	31	3.81
Anterior cingulate cortex	R	12	47	13	21	3.79
Superior frontal gyrus	L	-12	26	49	16	3.77
Middle frontal gyrus	L	-33	29	40	22	3.72
Insula	R	36	-4	10	14	3.63
Cerebellum	R	24	-55	-23	21	3.60
Superior medial frontal gyrus	L	-9	59	16	13	3.59
Hippocampus (partly covering)	L	-27	-19	-5	30	3.55
Caudate nucleus	R	9	20	4	15	3.51
Superior parietal lobe	L	-18	-43	40	11	3.51
Hippocampus	R	21	-25	-11	19	3.50
Middle frontal gyrus	L	-30	47	28	7	3.39
Dorsal premotor area	L	-18	2	61	11	3.38
Hippocampus	R	30	-43	-2	6	3.34
Superior frontal gyrus	L	-21	56	10	8	3.23

Note: The values shown are Montreal Neurological Institute (MNI) coordinates for significant activation maxima of clusters in the random effects analyses ($p < 0.001$, uncorrected). BA, Brodmann area; R, right; L, left.

Table 6
Brain activations during observation of finger movements corresponding to melodies learned in the audio-motor training condition vs. learned in the visuo-motor training condition (VpresAtrain vs. VpresVtrain, $p < .05$, FDR corrected, masked with motor execution map).

Anatomical region	Hemi-sphere	MNI coordinates			Cluster size	Z-score
		x	y	z		
IPL/somatosensory cortex	R	63	-22	40	181	4.55
IPL/somatosensory cortex	L	-63	-25	31	141	3.84
Superior parietal lobe	R	12	-70	46	21	3.77
Supplementary motor area/ Middle cingulate cortex	R/L	6	14	49	160	3.68
BA44	L	-57	5	34	94	3.60
Rolandic operculum		-48	2	13		
Dorsal premotor area (SFG/BA6)	L	-18	2	58	22	3.47
Dorsal premotor area (MFG/BA6)	R	33	-1	55	141	3.43
BA44	R	57	11	19	51	3.35
Brainstem	L/R	-3	-25	-2	31	3.31
BA44	R	51	11	-2	18	3.22
IPL	R	60	-34	28	13	2.87
Anterior insula	L	-30	26	4	8	2.75

Note: The values shown are Montreal Neurological Institute (MNI) coordinates for significant activation maxima of clusters in the random effects analyses ($p < 0.05$, FDR corrected). IPL, inferior parietal lobe; SFG, superior frontal gyrus; MFG middle frontal gyrus; BA, Brodmann area; R, right; L, left.

revealed that participants did not yet establish perfect tone–finger associations during the course of audio–motor training sessions, because often more trials and therefore a certain amount of time were needed to find the specific finger movement in association to a heard piano tone. Taken together, this might explain why the matching to a sequence of trained finger movements during listening to a melody was less successful.

The fMRI data had been collected *before* the recognition test and did not involve conscious recognition of trained sequences (the attention of participants was directed to the position of the played sequences rather than to the melody itself). To test for cross-modal transfer, we compared brain activity during the perception of trained versus untrained sequences, both presented in the other, non-training modality. All these sequences were perceptually unfamiliar for our participants: they had not previously been exposed to the untrained sequences, and had never heard corresponding melodies of the visuo-motor trained sequences or seen corresponding finger movements of the audio-motor trained sequences. Stronger brain activation for trained vs. untrained sequences would thus support the assumption for cross-modal transfer of information stemming from differences in familiarity in the other domains (auditory or motor domain for visual presentation and visual or motor domain for auditory presentations). The contrast “observation of finger movements corresponding to melodies trained in the audio-motor condition vs. observation of untrained finger movements” (VpresAtrain vs. VpresUntrain) revealed stronger activity in the rolandic operculum (at $p < .05$, FDR corrected) extending into BA44 (at a more liberal significance level, $p < .001$, uncorrected). These brain areas were also more strongly activated during the “execution of finger vs. eye movements” in the same participants (see also Fink et al., 1997). BA44 has been proposed to be a core area of an action-perception matching system that transforms perceived actions into internal representations (for review see Rizzolatti and Craighero, 2004). In accordance with this, BA44 has been reported to be more strongly activated during the observation of actions for which the observer possesses motor experience, such as the observation of actions of conspecifics (Buccino et al., 2004), observation of trained dance movements (Calvo-Merino et al., 2005, 2006; Cross et al., 2006) or observation of musical actions (Haslinger et al., 2005). Furthermore, studies investigating audio-motor associations and the effect of motor experience in the auditory domain found stronger activation in BA44 and adjacent brain regions for listening to trained piano melodies vs. untrained piano melodies (Lahav et al., 2007; Mutschler et al., 2007; see also Bangert et al., 2006 for a comparison of musicians vs. non-musicians during listening to piano melodies and Gazzola et al., 2006 for auditory action perception). We assume that the perception of finger movements elicited a co-activation of motor representations to reproduce these movements, which were consolidated in the audio-motor training sessions. More brain activation when viewing the audio-motor compared to the untrained sequences shows that not only motor representations corresponding to the individual notes were evoked: both types of melodies were composed of the same five notes, and would thus have caused the same motor activity only if motor representations corresponding to individual notes were activated during observation of the finger movements. Instead, higher-level representations of the sequential grouping of notes into melodies must have been recruited, as only this sequence level would have richer motor representations for the trained vs. untrained sequences. Interestingly, the rolandic-opercular area, which we found to be active most reliably in the VpresAtrain vs. VpresUntrain contrast, is often discussed as part of the vocal motor tract (see Brown et al., 2005) containing a representation of the larynx (see Koelsch et al., 2006), and seems to have impaired functioning in stutterers (Brown et al., 2005). Consistent with that, Schubotz and von Cramon (2003) show that the inferior part of the ventrolateral premotor cortex (including parts of the BA44) is related to mouth movements (see also Gazzola et al., 2006) and furthermore plays a role in rhythmic sequence recognition (Schubotz, 2007;

Schubotz and von Cramon, 2003). Our results support the assumption that cross-modal transfer in non-musicians may depend on motor regions associated with vocalization and that processes of sub-vocalization might be an important element in processing the sequential and rhythmic order of the movements. Consistent with this is a result on silent reading of music scores (which can be seen as analogous to observing silent finger movements that would potentially produce a melody on a piano keyboard), during which a process of sub-vocalization (facilitating “internally hearing”) was found in musicians by recording throat-audio and larynx-electromyography (Brodsky et al., 2008).

Moreover, a relation to the corresponding motor pattern during the observation of finger movements learned in the audio-motor training condition is backed by two further results: first, other motor-related areas, including the dorsal premotor cortex, cerebellum and putamen, were more active for observation of audio-motor trained vs. untrained sequences (VpresAtrain vs. VpresUntrain, at the more liberal significance level, $p < .001$, uncorrected). All these areas were also more strongly activated during the “execution of finger vs. eye movements”. The dorsal premotor cortex is involved in the control and the perception of hand actions (Gazzola et al., 2006, 2007). Putamen and cerebellum are involved in both motor control (Witt et al., 2008) and rhythm and beat perception (Grahn and Brett, 2007; Grahn and Rowe, 2009). Second, the results of the contrast “observation of finger movements corresponding to melodies trained in the audio-motor condition vs. observation of finger movements trained in the visuo-motor condition” (VpresAtrain vs. VpresVtrain) revealed brain activity in a network overlapping with that found in the contrast VpresAtrain vs. VpresUntrain (namely BA44, the left rolandic operculum and the dorsal premotor cortex) in addition to other motor-related areas. The vision of finger movements from the visuo-motor training condition is highly familiar to participants while the vision of those from the audio-motor training condition is not. Differences in brain activation between these conditions could thus in principle reflect perceptual familiarity, with the less familiar sequences (VpresAtrain) requiring more attention and hence more brain activity. However, the finding that some of the same brain regions respond more to VpresAtrain than VpresUntrain, where perceptual familiarity is low in both cases, makes this interpretation of perceptual familiarity unlikely. Thus, a more parsimonious interpretation for the overlapping results of these two contrasts is that VpresAtrain sequences caused the largest activation because it requires more cross-modal transfer than VpresVtrain sequences (where presentation and training procedure were in the same modality) and triggers richer motor representations than VpresUntrain sequences.

Taken together, these results, by suggesting a co-activation of related motor patterns during the observation of finger movements corresponding to melodies trained in the audio-motor condition, provide evidence for the motor mediation hypothesis of cross-modal transfer (as outlined in the Introduction) from the audio-motor training condition to the visual domain. In addition to, or instead of, differences in motor-related brain activation, trained compared to untrained sequences might also have elicited stronger activations in the sensory modality in which the sequences had been trained. Such a result would rather speak for a sensory mediation of cross-modal transfer from the audio-motor training condition to visual perception (see sensory mediation hypothesis in the introduction). Studies in musicians who trained piano pieces under natural conditions that involved audition, vision and motor/kinesthetic modalities have shown such effects, with observation of (silent) finger movements triggering activations in auditory cortices (Hasegawa et al., 2004; Haslinger et al., 2005). In our data, however, a whole brain analysis, even at a more liberal significance level ($p < 0.001$, uncorrected), did not show auditory sensory areas more strongly activated during “observation of blindly audio-motor trained vs. untrained movements”. Please note that testing at a more liberal significance level increases the chances of detecting evidence against the preferred motor mediation hypothesis.

Our data therefore warrant the conclusion that a transfer of information from the audio-motor training condition to the visual testing domain did occur, and that this transfer seems to have relied more on the motor than the sensory content of the original training condition and therefore supports the motor mediation hypothesis. Furthermore, our results suggest that this transfer may occur obligatorily, since the task during perception was to discriminate the *position* of the sequences and not whether it had been trained. However, with this study it remains unclear whether sensory modalities might play a more crucial role in cross-modal transfer for untrained sequences in musicians, who trained over long periods of time all three modalities together (visual, auditory and motor domain; c.f. Hasegawa et al., 2004; Haslinger et al., 2005).

The contrast for the auditory presentation comparing sequences trained in the silent visuo-motor condition vs. untrained sequences (ApresVtrain vs. ApresUntrain) did not reveal any significant differences. Thus, in contrast to the content of audio-motor training condition that becomes available to the visual modality, even after three days of visuo-motor and three days of audio-motor training sessions, the content of the visuo-motor training condition remained inaccessible to the auditory modality. Given limited statistical power, this does not exclude the presence of some cross-modal transfer from the visuo-motor to the auditory domain, but indicates that transfer seemed strong enough to be significant only for the audio-motor to the visual domain. Studies of musicians, however, show stronger brain activation in motor-related areas during listening to trained melodies (e.g. Bangert et al., 2006), and even non-musicians show stronger brain activity after a short audio-visuo-motor training procedure in motor-related areas for trained compared to untrained sequences, which holds true when auditory familiarity is controlled (Lahav et al., 2007; Mutschler et al., 2007). The critical difference with our study is that we trained the non-musicians motorically in connection with one modality and we test them in the other, while participants in previous experiments had been trained without excluding the testing modality. Thus, for a transfer of an auditory sequence into a motor pattern to occur in *non-musicians*, participants seem to have to be trained on these specific sequences using the auditory modality. Or, as an alternative, they must have received a much more extensive audio-motor training procedure. This training scheme could have established such strong audio-motor connections between heard single tones and corresponding finger movements, that could be accessed during listening to an unfamiliar melody – and used to associate a whole sequence of a motor program. The process of practicing other sequences in the audio-motor training condition seems to have generated some associations between individual notes and finger movements, as evidenced by the participants' performance at the test for audio-motor associations that required finding the corresponding finger movements for perceived tones. However, these tone–finger associations were not that reliable yet, as often more trials were needed to reproduce a heard piano tone. The audio-motor training sessions did not aim to establish note–finger associations explicitly, but rather to train the entire sequence of tones (i.e., the melody). After remembering a specific sequence, a mapping between specific tones and finger movements was not necessary anymore for successful performance. Taken together, tone–finger associations acquired during the audio-motor training sessions seem to have been too weak to specifically trigger melody representations significantly, and hence more motor activation when listening to visuo-motor trained sequences than untrained melodies (c.f., Eldridge et al., 2010 who showed poorer pitch-motor mapping after audio-motor compared to audio-visuo-motor training procedure of piano pieces).

Conclusions

In this study we trained non-musicians to play sequences on a piano keyboard by either merely listening and replaying (vision of own fingers occluded) or merely observing silent finger movements

and replaying (on a silent keyboard). We show that participants were more accurate at recognizing the sight of finger movements corresponding to the melodies that had been learned in the audio-motor training condition than the sound of melodies corresponding to the sequences that had been trained in the visuo-motor condition. Observing finger movements corresponding to audio-motor trained melodies elicited stronger reliable brain activity in the rolandic operculum – a brain area which was also more strongly activated during the execution of finger movements – than observing untrained sequences. Taken together, behavioral results demonstrate a transfer of information from the audio-motor training domain to the visual presentation domain. The neuroimaging findings suggest that this transfer is facilitated by the activation of motor sequence programs during the observation of blindly trained finger movements. No such neural evidence was found for cross-modal transfer for the visuo-motor training condition to the auditory presentation suggesting that the later might be reached only by more extensive training schemes. All together these results support the motor mediation hypothesis of cross-modal transfer from the audio-motor training condition to the visual domain.

Author contributions

AE and CK designed the fMRI experiment; AE, MB, PK, DH and CK designed the training procedure and stimuli; AE, BH, DH, and KW acquired the data of the reported and a preceding pilot experiment; AE analyzed the data with advices from and consultations with CK, MB and PK; the manuscript was written by AE with contributions from all authors.

Supplementary data to this article can be found online at [doi:10.1016/j.neuroimage.2012.03.038](https://doi.org/10.1016/j.neuroimage.2012.03.038).

Acknowledgments

This research was supported by a Postdoc grant of the German Research Foundation (DFG) assigned to AE (EN897/1-1; EN897/1-2), by a Marie Curie Excellent Grant of the European commission (MEXT-CT-2005-023253), and a N.W.O. VIDI (452-04-305) grant to CK and the Max Planck Society through the Music Cognition and Action group of PK. We thank Luca Nanetti, Anita Kuiper and Judith Streurman for scanning the fMRI experiment, and our participants for their patience in taking part in extensive training and scanning procedures.

References

- Bangert, M., Altenmüller, E., 2003. Mapping perception to action in piano practice: a longitudinal DC-EEG study. *BMC Neurosci.* 4, 26.
- Bangert, M., Häusler, U., Altenmüller, E., 2001. On practice: how the brain connects piano keys and piano sounds. *Ann. N. Y. Acad. Sci.* 930, 425–428.
- Bangert, M., Peschel, T., Schlaug, G., Rotte, M., Drescher, D., Hinrichs, H., Heinze, H.J., Altenmüller, E., 2006. Shared networks for auditory and motor processing in professional pianists: evidence from fMRI conjunction. *NeuroImage* 30, 917–926.
- Brass, M., Bekkering, H., Wohlschläger, A., Prinz, W., 2000. Compatibility between observed and executed finger movements: comparing symbolic, spatial, and imitative cues. *Brain Cogn.* 44, 124–143.
- Brass, M., Bekkering, H., Prinz, W., 2001. Movement observation affects movement execution in a simple response task. *Acta Psychol. (Amst)* 106, 3–22.
- Brodsky, W., Kessler, Y., Rubinstein, B.S., Ginsborg, J., Henik, A., 2008. The mental representation of music notation: notational audiation. *J. Exp. Psychol. Hum. Percept. Perform.* 34, 427–445.
- Brown, S., Ingham, R.J., Ingham, J.C., Laird, A.R., Fox, P.T., 2005. Stuttered and fluent speech production: an ALE meta-analysis of functional neuroimaging studies. *Hum. Brain Mapp.* 25, 105–117.
- Buccino, G., Lui, F., Canessa, N., Pastteri, I., Lagravinese, G., Benuzzi, F., Porro, C.A., Rizzolatti, G., 2004. Neural circuits involved in the recognition of actions performed by nonconspecifics: An fMRI study. *J. Cogn. Neurosci.* 16, 114–126.
- Bushnell, E.W., Baxt, C., 1999. Children's haptic and cross-modal recognition with familiar and unfamiliar objects. *J. Exp. Psychol. Hum. Percept. Perform.* 25, 1867–1881.
- Calvo-Merino, B., Glaser, D.E., Grezes, 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., Grezes, 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.
- Casile, A., Giese, M.A., 2006. Nonvisual motor training influences biological motion perception. *Curr. Biol.* 16, 69–74.
- Cross, E.S., Hamilton, A.F., Grafton, S.T., 2006. Building a motor simulation de novo: observation of dance by dancers. *NeuroImage* 31, 1257–1267.
- Davis, M.H., 1983. Measuring individual differences in empathy: evidence for a multi-dimensional approach. *J. Pers. Soc. Psychol.* 44, 113–126.
- Del Giudice, M., Manera, V., Keyzers, C., 2009. Programmed to learn? The ontogeny of mirror neurons. *Dev. Sci.* 12, 350–363.
- Eldrige, M., Saltzman, E., Lahav, A., 2010. Seeing what you hear: visual feedback improves pitch recognition. *J. Cogn. Psychol.* 22, 1078–1091.
- Fink, G.R., Frackowiak, R.S., Pietrzyk, U., Passingham, R.E., 1997. Multiple nonprimary motor areas in the human cortex. *J. Neurophysiol.* 77, 2164–2174.
- Fleming, N.D., Mills, C., 1992. Not another inventory, rather a catalyst for reflection. *To Improve the Academy* 11, 137.
- Gazzola, V., Aziz-Zadeh, L., Keyzers, C., 2006. Empathy and the somatotopic auditory mirror system in humans. *Curr. Biol.* 16, 1824–1829.
- Gazzola, V., Rizzolatti, G., Wicker, B., Keyzers, C., 2007. The anthropomorphic brain: the mirror neuron system responds to human and robotic actions. *NeuroImage* 35, 1674–1684.
- Grahn, J.A., Brett, M., 2007. Rhythm and beat perception in motor areas of the brain. *J. Cogn. Neurosci.* 19, 893–906.
- Grahn, J.A., Rowe, J.B., 2009. Feeling the beat: premotor and striatal interactions in musicians and nonmusicians during beat perception. *J. Neurosci.* 29, 7540–7548.
- Halpern, A.R., Zatorre, R.J., 1999. When that tune runs through your head: a PET investigation of auditory imagery for familiar melodies. *Cereb. Cortex* 9, 697–704.
- Hasegawa, T., Matsuki, K., Ueno, T., Maeda, Y., Matsue, Y., Konishi, Y., Sadato, N., 2004. Learned audio-visual cross-modal associations in observed piano playing activate the left planum temporale. An fMRI study. *Brain Res. Cogn. Brain Res.* 20, 510–518.
- Haslinger, B., Erhard, P., Altenmüller, 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.
- Haueisen, J., Knösche, T.R., 2001. Involuntary motor activity in pianists evoked by music perception. *J. Cogn. Neurosci.* 13, 786–792.
- Hecht, H., Vogt, S., Prinz, W., 2001. Motor learning enhances perceptual judgment: a case for action–perception transfer. *Psychol. Res.* 65, 3–14.
- Hommel, B., Müsseler, J., Aschersleben, G., Prinz, W., 2001. The theory of event coding (TEC): a framework for perception and action planning. *Behav. Brain Sci.* 24, 849–878 discussion 878–937.
- Iacoboni, M., Woods, R.P., Brass, M., Bekkering, H., Mazziotta, J.C., Rizzolatti, G., 1999. Cortical mechanisms of human imitation. *Science* 286, 2526–2528.
- Kato, M., Konishi, Y., 2006. Auditory dominance in the error correction process: a synchronized tapping study. *Brain Res.* 1084, 115–122.
- Keyzers, C., Perrett, D.I., 2004. Demystifying social cognition: a Hebbian perspective. *Trends Cogn. Sci.* 8, 501–507.
- Koelsch, S., Fritz, T., Cramon, D.Y.V., Müller, K., Friederici, A.D., 2006. Investigating emotion with music: an fMRI study. *Hum. Brain Mapp.* 27, 239–250.
- Lacey, S., Campbell, C., 2006. Mental representation in visual/haptic crossmodal memory: evidence from interference effects. *Q. J. Exp. Psychol. (Colchester)* 59, 361–376.
- 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.
- Lotze, M., Scheler, G., Tan, H.R., Braun, C., Birbaumer, N., 2003. The musician's brain: functional imaging of amateurs and professionals during performance and imagery. *NeuroImage* 20, 1817–1829.
- Mutschler, I., Schulze-Bonhage, A., Glauche, V., Demandt, E., Speck, O., Ball, T., 2007. A rapid sound–action association effect in human insular cortex. *PLoS One* 2, e259.
- Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113.
- Platel, H., Price, C., Baron, J.C., Wise, R., Lambert, J., Frackowiak, R.S., Lechevalier, B., Eustache, F., 1997. The structural components of music perception. A functional anatomical study. *Brain* 120 (Pt 2), 229–243.
- Reithler, J., van Mier, H.L., Peters, J.C., Goebel, R., 2007. Nonvisual motor learning influences abstract action observation. *Curr. Biol.* 17, 1201–1207.
- Repp, B.H., Penel, A., 2002. Auditory dominance in temporal processing: new evidence from synchronization with simultaneous visual and auditory sequences. *J. Exp. Psychol. Hum. Percept. Perform.* 28, 1085–1099.
- Repp, B.H., Penel, A., 2004. Rhythmic movement is attracted more strongly to auditory than to visual rhythms. *Psychol. Res.* 68, 252–270.
- Rizzolatti, G., Craighero, L., 2004. The mirror-neuron system. *Annu. Rev. Neurosci.* 27, 169–192.
- Roland, P.E., Gulyas, B., 1994. Visual imagery and visual representation. *Trends Neurosci.* 17, 281–287 discussion 294–287.
- Rossion, B., Schiltz, C., Robaye, L., Pirenne, D., Crommelinck, M., 2001. How does the brain discriminate familiar and unfamiliar faces?: A PET study of face categorical perception. *J. Cogn. Neurosci.* 13, 1019–1034.
- Schubotz, R.I., 2007. Prediction of external events with our motor system: towards a new framework. *Trends Cogn. Sci.* 11, 211–218.
- Schubotz, R.I., von Cramon, D.Y., 2003. Functional-anatomical concepts of human premotor cortex: evidence from fMRI and PET studies. *NeuroImage* 20 (Suppl. 1), S120–S131.
- Witt, S.T., Laird, A.R., Meyerand, M.E., 2008. Functional neuroimaging correlates of finger-tapping task variations: an ALE meta-analysis. *NeuroImage* 42, 343–356.
- Zatorre, R.J., Halpern, A.R., 2005. Mental concerts: musical imagery and auditory cortex. *Neuron* 47, 9–12.