Neural networks for harmonic structure in music perception and action

Bianco R. [1], Novembre G.[2], Keller, P.E.[2], Kim SG[1], Scharf, F.[1], Friederici A.D.[1], Villringer A.[1], Sammler D.[1]

[1] Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

[2] The MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Sydney, Australia

Address correspondence to:
Roberta Bianco,
Max Planck Institute for Human Cognitive and Brain Sciences
Stephanstr. 1a, 04103 Leipzig, Germany
phone: +49 341 9940 2640
Email: bianco@cbs.mpg.de

Abstract

The ability to predict upcoming structured events based on long-term knowledge and contextual priors is a fundamental principle of human cognition. Tonal music triggers predictive processes based on structural properties of harmony, i.e., regularities defining the arrangement of chords into well-formed musical sequences. While the neural architecture of structure-based predictions during music *perception* is well described, little is known about the neural networks for analogous predictions in musical *actions* and how they relate to auditory perception. To fill this gap, expert pianists were presented with harmonically congruent or incongruent chord progressions, either as musical actions (photos of a hand playing chords) that they were required to watch and imitate without sound, or in an auditory format that they listened to without playing. By combining task-based functional magnetic resonance imaging (fMRI) with functional connectivity at rest, we identified distinct sub-regions in right inferior frontal gyrus (rIFG) interconnected with parietal and temporal areas for processing action and audio sequences, respectively. We argue that the differential contribution of parietal and temporal areas is tied to motoric and auditory long-term representations of harmonic regularities that dynamically interact with computations in rIFG. Parsing of the structural dependencies in rIFG is co-determined by both stimulus- or task-demands. In line with contemporary models of prefrontal cortex organization and dual stream models of visual-spatial and auditory processing, we show that the processing of musical harmony is a network capacity with dissociated dorsal and ventral motor and auditory circuits, which both provide the infrastructure for predictive mechanisms optimising action and perception performance.

Keywords: Music, harmony, syntax, IFG, functional connectivity, prediction.

1. Introduction

The brain shows a fine sensitivity to patterns and regularities that afford the prediction of incoming events in different domains (Tenenbaum et al., 2011). The theory of predictive coding (Friston, 2010) constitutes a unifying framework for human cognition and considers the brain as a "hypothesis tester" with the goal to optimise perception and action by constantly matching incoming sensory inputs with top-down predictions. Within a multi-level cascade of neural processes at different time scales, higher-level predictions act as priors for lower-level processes based on contextual information, previous exposure and acquired long-term knowledge. Recently, predictive coding theory has been used to explain predictions in the action domain (Kilner et al., 2007), as well as in music perception based on priors related to melodic (pitch) content (Pearce et al., 2010), metric structure (Vuust and Witek, 2014), or harmony (Rohrmeier and Koelsch, 2012). The present study takes a comparative stance on predictions in both music perception and action, with a specific focus on Western tonal harmony.

Theoretical accounts refer to harmony as combinatorial arrangement of chords within musical sequences characterized by local and non-local dependencies (Swain, 1995). An instance of these dependencies is that a typical chord progression in Western tonal harmony starts and ends with a reference chord to which some chords are overwhelmingly likely to move to, while they rarely move to others (Tymoczko, 2003). Psychologically, these dependencies are predicted and perceived as tension-resolution patterns by listeners who have been sufficiently exposed to the prevailing musical system (Krumhansl, 1983; Lerdahl and Jackendoff, 1983). Convention in the field of music cognition has that the harmonic principles that govern musical structure are considered as part of a musical "syntax" (Bharucha and Krumhansl, 1983; Koelsch and Siebel, 2005; Patel, 2003), that also includes melodic and/or rhythmic principles of music (Large and Palmer, 2002; Rohrmeier and Koelsch, 2012). Here, we consider "syntax" generally as the knowledge of regularities that control the integration of smaller units into larger musical phrases (Swain, 1995) and thereby support predictions. It is well established that tacit knowledge about structural regularities of music 1) is acquired implicitly (Loui et al., 2009; Rohrmeier and Rebuschat, 2012; Tillmann et al., 2000), 2) largely shapes our musical competence across different musical systems and cultures (Eerola et al., 2006; Lartillot and Ayari, 2011), and 3) enables listeners to cognitively link current auditory items to past events and to generate predictions on forthcoming events (Patel, 2003; Tillmann, 2012). In the present study we will focus on harmonic regularities and investigate how they govern predictions during (auditory) music perception and (silent) musical actions.

Harmony not only defines the sequence of musical *sounds* but also co-determines the associated chain of musical *actions*. Therefore, the implicit knowledge of harmonic regularities might influence not only *listeners'* predictions, but also musicians' *action* planning during performance (Palmer and van de Sande, 1995). While regularity-based

predictions during music *listening* have already been thoroughly investigated (Rohrmeier and Koelsch, 2012; Tillmann, 2012), the neural basis of motor predictions in musical *actions* has not been explored in depth (Maidhof et al., 2009; Ruiz et al., 2009). Recent behavioural (Novembre and Keller, 2011) and electrophysiological studies on music production (Bianco et al., 2016; Sammler et al., 2013b) revealed slower response times, higher number of errors and neural processing costs (a centro-parietal negativity) in expert pianists when asked to silently execute harmonically incongruent compared to congruent chord progressions. These costs were associated with the motor reprogramming of a pre-planned, congruent, action in face of an unexpected incongruity, and were taken as indirect evidence that pianists' action planning was based on musical context and internalised knowledge of harmony. In other words, these findings imply that harmonic structure might implicitly regulate mechanisms of motor control to improve music performance beyond fine movement optimization (Bianco et al., 2016; Novembre and Keller, 2011).

The goal of the present study is to identify the brain areas involved in motor planning based on the regularities of Western tonal harmony, to explore the connectivity between these areas and to compare this network with the neural network sub-serving analogous processes in auditory music perception. The rationale behind this study is that expert pianists have internalized the rules of harmony not only auditorily but also in the hand action domain. Their substantial motor training should enable them to parse harmonic dependencies also in sequences of silent musical actions to facilitate prediction and planning of forthcoming motor acts during performance. This is because the same harmonic structure in sequences of sounds or sequences of actions without sound (i.e those movements typically employed for producing these sounds) should trigger cognitive processes that are analogous with regard to the structural information. At the same time, processing should differ between perception and action with regard to the associated sensory and memory retrieval processes (i.e., auditory sound vs. motoric act). Here, we sought to isolate and compare the neural networks involved in harmony processing during either perception or (silent) actions, i.e. to probe the potential contribution of auditory and motor prediction of harmony that are otherwise co-occuring during real music production.

Neural hypotheses for musical syntax processing (i.e., harmony) in music *perception* (Koelsch, 2011; Patel, 2003; Tillmann, 2012) posited a special role of frontal computational regions that successively integrate incoming information into higher-order structures by drawing on knowledge about regularities stored in posterior brain regions. Neuroimaging research points to the inferior frontal gyrus (IFG) as the critical computational area that, together with a repository of regularities in posterior auditory regions superior temporal gyrus (STG), affords the prediction of future musical sounds based on the context and listener's long-term music structural knowledge (Kim et al., 2011; Koelsch et al., 2005; Maess et al., 2001; Musso et al., 2015; Sammler et al., 2011; Tillmann et al., 2006). Interestingly, IFG has been associated not only with structural integration and prediction of musical sequences, but

also with structuring of complex *actions* (Fuster, 2001; Koechlin and Summerfield, 2007) outside the music domain. Lesions of the left IFG cause impairment in sequencing pictures representing human actions (Fazio et al., 2009), and bilateral IFG are involved in evaluating whether constituent acts belong to the same or separate sub-goals (Farag et al., 2010). Moreover, bilateral IFG activations have been reported during execution of series of motor acts that were organised according to hierarchical action plans (Koechlin and Jubault, 2006). In sum, IFG has become central to hypotheses on processing of structured sequential information in perception and action (Fitch & Martins, 2014; Fiebach & Schubotz, 2006, for various perspectives see Cortex, 2006, vol.2, issue 42), making it conceivable that IFG is also involved in parsing and predicting structural information embedded in musical actions.

What has received less attention than the role of IFG, however, is its interaction with task-relevant posterior systems of knowledge during structural processing. In other words, apart from frequently reported co-activations of IFG and auditory temporal regions during music listening (Koelsch and Siebel, 2005), the characterization of other 'modality-specific regions', e.g., in musical action, and particularly their connectivity with frontal 'computational regions' remains uncertain. In this study, we tested whether pianists' action planning based on knowledge of Western tonal harmony involves (i) IFG in interaction with (ii) posterior visual-motor areas. Furthermore, we (iii) compared the functional connectivity profiles of IFG during the processing of musical actions and auditory sequences that contained similar harmonic violations.

We acquired resting state fMRI data from expert pianists, and then fMRI data during an audio and an action task in which the same harmonic sequences were either auditorily presented or had to be motorically imitated. In the audio task, pianists listened to 5-chord sequences (similar to Koelsch et al., 2005) in which the last chord was either harmonically congruent or incongruent with the preceding musical context. In the action task, in total absence of musical sound, participants were presented with series of photos of a pianist's hand performing the same congruent/incongruent chord progressions on a piano (Bianco et al., 2016). To engage the motor system in the processing of musical actions, pianists had not only to watch the movements, but also to manually reproduce them on a glass-board. The contrasts of incongruent minus congruent chords during listening or imitation were used to functionally segregate modality-specific areas and to isolate frontal computational areas. To demonstrate crosstalk between these regions, we used the latter as seeds in a functional connectivity analysis of the resting state fMRI data.

If harmonic violations of audio sequences activate IFG, then violations of action sequences with the same musical structure should also activate IFG as parser of harmonic regularities and top-down generator of predictions. On the other hand, we expected to find divergent activity in temporal auditory or parietal visual-motor regions associated with item identification and storage of knowledge in their modality-specific format. Finally, by mirroring task-based activation (Smith et al., 2009), the resting-state data should reveal

processing streams involved in processing harmonic regularities in music perception and action.

2. Materials and methods

2.1 Participants.

29 pianists (17 female) aged 20-32 years (mean age: 24.7, SD = 2.9) took part in the experiment. They had a minimum of 5 years of piano training in classical Western tonal music (range = 5 - 27 years, mean years of training = 17.2, SD = 4.8) and had started to play the piano at an average age of 7.3 years (SD = 3.08). None of the pianists had training in improvisation or other musical styles. All participants were naïve with regard to the purpose of the study. Written informed consent was obtained from each participant before the study that was approved by the local ethics committee.

2.2 Stimuli.

Stimuli (see Fig. 1) consisted of 60 different chord sequences that were presented as piano sounds in the listening task (similar to Koelsch, 2005), and as photos of a hand playing chords on a piano in the action imitation task (Bianco et al., 2016). The sequences were composed of 5 chords according to the rules of classical harmony and had various melodic contours. The first chord always represented the tonic (based on the first degree of the scale in the relevant musical key). The second chord could be tonic, mediant (based on the third scale degree) or subdominant (based on the fourth scale degree). Chords at the third position were subdominant, dominant, or dominant six-four chords, and chords at the fourth position were dominant seventh chords. At the last position, the target chord of each sequence was manipulated in terms of harmonic congruency (CONG), so that the last chord could be either congruent (a Tonic chord typically used to resolve a musical sequence) or incongruent (a Neapolitan chord that sounds normal when played in isolation but constitutes a violation when used at the end of a standard harmonic progression). Both the Tonic and Neapolitan are consonant major chords built on the 1st and lowered 2nd scale degree, respectively (i.e., A for Tonic and Bb for Neapolitan in A-major). Consequently, and due to the relationship of the tonalities within the circle of fifths, the exact same chord that acts as a Tonic in one tonality (e.g., A – #C – E in A-major), acts as a Neapolitan in another tonality (i.e., Bbb (=A) – Db (= #C) – Fb (= E) in Ab-major). We exploited this relationship and presented five sequences from each of six different tonalities (D, E, Bb, Ab, A and Eb major), such that the majority of final chords were presented as both Tonic and Neapolitan across the experiment. Therefore, potential neural differences in processing congruent and incongruent chords cannot be due to chord identity but more likely reflect harmony-related processes. With regard to the stimuli of the action block, the choice of different tonalities further allowed us to balance the visual appearance (i.e., number of black/white keys) and difficulty of execution of the target chord (i.e., movement distance from second last to target chord) in congruent and incongruent conditions (for visual appearance: average of 1.3 ± 0.5 black keys in the congruent and 1.2 ±

0.7 in the incongruent chords; for difficulty: average of 1± 0 key distance in the congruent and 1.5 ± 0.2 in the incongruent sequences). All sequences were played with normal fingering that was rated as being similarly conventional for congruent and incongruent endings (see Bianco et al., 2016).

The audio stimuli were created with Logic Pro 8 (Apple Inc.), normalised for loudness (RMS, root mean square) with Adobe Audition CS 6 and had a total duration of 6 seconds (1 second for each of the first four chords and 2 seconds for the target chord). In the action block, the same chord sequences were presented as photo series showing a male pianist's right hand pressing three keys forming each chord on a piano in conventional fingering (Yamaha Clavinova CLP150, Yamaha Music Europe GmbH, Rellingen, Germany). Red circles were superimposed on top of each pressed key (cf. Bianco et al, 2016) for the whole duration of the photo to facilitate the recognition of the pressed keys. Each photo was presented for 2 seconds (total sequence duration: 10 seconds).

2.3 Procedure.

The experimental session started with 14 minutes resting state fMRI data acquisition in which participants were instructed to keep their eyes open and not to fall asleep. To prevent any task-related bias in the measures of functional connectivity, pianists were asked not to practice piano on the scanning day. Thereafter, the task session started and lasted for approximately 25 minutes.

In the scanner, participants were required to imitate musical actions or to listen to musical sequences in two separate blocks with a counterbalanced order across the group (Fig.1). Stimulus presentation was controlled in an event-related design with Presentation software (version 14.9, Neurobehavioural Systems, Inc.). In both blocks, congruent and incongruent sequences were intermixed in a way that no more than 3 sequences of the same condition followed each other. The inter-trial interval (ITI) ranged from 3 to 9 seconds and during this period participants saw a black screen. During action imitation, no sound was played. Pianists were asked to watch the performing hand in the photos and to simultaneously copy the presented hand postures on a 5x15 cm^2 glass-board with their right hand (Fig.1, left panel). To motivate participants to follow the sequence accurately, they were told that their performance was monitored with a camera (MR-compatible camera, 12M camera, MRC Systems, Heidelberg Germany). Since it is a common way for pianists to mentally practice by motorically simulating piano performance, the playing along was meant to maximally involve the motor system during the processing of musical actions. In the audio block, pianists were asked to carefully listen to the sequences without playing along (Fig.1, right panel).

Only to ensure that participants paid attention to the stimuli and to assess their awareness of the violations, 10 trials (1/6 of the trials) in both tasks were followed by a prompt that asked participants to judge the harmonic correctness of the last presented sequence. The judgement required a button response performed with the index or middle finger of the left hand (key assignment was counterbalanced across participants). These

judgement trials were equally distributed over congruent and incongruent trials, and required 50% yes and 50% no responses.

After the scanning session, participants filled out a questionnaire to assess the degree to which they had imagined the sound of the chord sequences during the action imitation task and the movements to produce the sequences during the listening task.

Figure 1

Figure 1. Experimental design: expert pianists were presented with harmonically congruent or incongruent chord progressions, presented either as muted musical actions (photos of a hand playing chords) that they were required to imitate on a glass-board (left panel), or in an auditory format that they listened to (right panel).

2.4 Data acquisition.

The experiment was carried out on a 3.0-Tesla Siemens TIM Trio whole body magnetic resonance scanner (Siemens AG, Erlangen, Germany) using a 32-radiofrequency-channel head coil. Functional magnetic resonance images were acquired using a T2*-weighted 2D echo planar imaging (EPI) sequence. During 14 minutes of acquisition (TE = 36.5 ms, TR = 1400 ms) at rest (eyes open, instructed not to fall asleep) 410 volumes were acquired with a square FOV of 64 axial slices of 2.3 mm thickness and no gap (2.3 x 2.3 x 2.3 mm^3 voxel size) with a flip angle of 69°. Functional images during the two tasks were acquired using an EPI sequence with TE = 30 ms and TR = 2000 ms. 456 and 377 volumes were acquired in the action and audio block, respectively, with a square FOV of 210 mm, with 37 interleaved slices of 3.2 mm thickness and 15% gap (3 x 3 x 3.68 mm^3 voxel size) aligned to the AC-PC plane, and a flip angle of 77°. For anatomical registration, high-resolution T1-weighted images were acquired using a 3D MP2RAGE sequence (TI_1 = 700 ms, TI_2 = 2500 ms, TE = 2.03 ms, TR = 5000 ms) with a matrix size of 240 x 256 x 176, with 1 mm isotropic voxel size, flip $angle_1$ of 4°, flip $angle_2$ of 8°, and GRAPPA acceleration factor of 3.

2.5 Data analysis.

2.5.1 Task-based fMRI

fMRI data of 29 participants were analysed with statistical parametric mapping (SPM8; Welcome Trust Centre for Neuroimaging; http://www.fil.ion.ucl.ac.uk/spm/software/spm8/) using standard spatial pre-processing procedures. These consisted of: slice time correction (by means of cubic spline interpolation method), spatial realignment, co-registration of functional and anatomical data (uniform tissue-contrast image masked with the 2nd inversion image from the MP2RAGE sequence), spatial normalisation into the MNI (Montreal Neurological Institute) stereotactic space, that included resampling to 2x2x2 mm voxel size. Finally, data were spatially low-pass filtered using a 3D Gaussian kernel with full-width at half-maximum

(FWHM) of 8 mm and temporally high-pass filtered with a cut-off of 1/128 Hz to eliminate low-frequency drifts.

Statistical parametric maps for the whole brain data were generated in the context of the general linear model (GLM) separately for the action imitation and the listening task. The evoked hemodynamic response to the onset of the final chord was modelled for the congruent and incongruent conditions as boxcars convolved with a hemodynamic response function (HRF). To this design, we added estimated motion realignment parameters as covariates of no interest to regress out residual motion artefacts and increase statistical sensitivity. To identify hemodynamic responses related to the processing of harmonic violations, we computed the first level contrast CONG (i.e., incongruent > congruent chords), separately for the action imitation and the listening task. For random effects group analyses, the resulting contrast images were submitted to one-sample *t*-tests. Additionally, to identify areas that are modality-specific to either action or audio representation of the harmonic structure, we compared the CONG contrasts of the two tasks by means of paired *t*-tests. We controlled family-wise error rate (FWER) of clusters below 0.05 with a cluster-forming height-threshold of 0.001. Anatomical labels are based on Harvard-Oxford cortical structural atlas implemented in FSL (http://neuro.debian.net/pkgs/fsl-harvard-oxford-atlases.html).

2.5.2 Resting-state fMRI

In order to investigate intrinsic connectivity of the peak regions from task-based fMRI datasets (Bressler and Menon, 2010), independent resting state fMRI datasets were obtained from 28 of the pianists that participated in the task-fMRI session (one r-fMRI data set was not acquired due to technical problems). The pre-processing of the resting state data (realignment, unwarping, slice-timing correction) was done using SPM8 by means of DPARSF (http://rfmri.org/DPARSF) SPM-based toolboxes. We applied a GLM to regress out non-neuronal signal changes due to physiological noise and, most importantly, head motions. The regressors included six rigid-body motion parameters, five principle components extracted by the "anatomical CompCor" (Behzadi et al., 2007) (i.e., signal from white matter and cerebral fluid masks defined from anatomical scans), and finally global signal (Power et al., 2015). Thereafter, band-pass-filtering (0.009 and 0.08 Hz), spatial normalization of functional data into MNI stereotactic space (with resampling to 2 x 2 x 2 mm^3 resolution), and finally a minimal spatial smoothing with the FWHM of 3 mm were applied to the residual time-series.

Resting-state functional connectivity (RSFC) was defined by Pearson's correlations between a time-series of a seed region and time-series of whole brain voxels. Spherical seed regions (5 mm radius) were centred in IFG at the peak coordinates of the CONG contrasts obtained in the task-based analyses of the action imitation and the listening task. In order to match the smoothness of noise in task-based and resting-state analyses, the correlation maps were further smoothed with the FWHM of 2 mm, resulting in an effective FWHM of about 8 mm.

Voxel-wise paired t-tests were performed to identify differences between the two seed-based correlation maps (i.e., action and audio seeds). The normality assumption based upon the difference between the two correlation coefficients across subjects was fulfilled, as confirmed by Kolmogorov-Smirnov tests. We controlled FWER of clusters below 0.05 with a cluster-forming height-threshold of 0.001 in all reported results. Harvard-Oxford cortical structural atlas was used to assign anatomical labels.

3. Results.

3.1 Behavioural.

To ensure that participants paid attention to the stimuli in both modalities and that they were generally able to recognise the harmonic structure underlying the sequences, they were required to overtly judge harmonic congruency in 1/6 of the trials. They performed significantly above chance level in these explicit judgments both in the action (mean ± *SD*: 68.96 ± 27.06% correct, $p < .001$) and in the audio block (mean ± *SD:* 91.03 ± 16.40% correct, $p < .001$), as tested with one-sample *t*-tests against 50% chance level. Action block performance was lower than audio block performance ($t(28)= 4.704$, $p < .001$), partly due to a response bias towards "congruent" answers in the action block (32% of incongruent trials misclassified as congruent vs. 15% of congruent trials misclassified as incongruent: $t(28)=-2.95$, $p = 0.007$). These differences in explicit judgment might indicate that pianists were less consciously aware of the harmonic violations during action imitation than during passive listening, possibly because it is more taxing to copy sequences of actions on-line than to just listen. This may have led pianists to focus on the motor-executive task rather than harmonic relationships in the action block (see Discussion).

3.2 fMRI.

3.2.1 Fronto-parietal vs. fronto-temporal areas for musical action vs. perception.

In the action imitation task, the CONG contrast (incongruent vs. congruent chords) yielded larger hemodynamic responses in frontal and parietal areas, comprising the dorsal portion of rIFG (BA44) bordering precentral sulcus, and bilateral clusters extending from superior parietal cortex (SPL: BA7) to the inferior parietal and middle occipital gyrus (MOG: BA19) (Fig. 2 left-upper panel, Table 1).

In the listening task, the same contrast evoked stronger activity in frontal and temporal areas, including right IFG (BA44/45, peak in ventral BA45) and the right posterior superior temporal gyrus and sulcus (pSTG/STS: BA22) (Fig. 2 right-upper panel, Table 1).

Figure 2

Figure 2. Harmonic violations elicited activations in fronto-parietal areas during action imitation (upper left panel) and in fronto-temporal areas during listening (upper right panel). Areas involved in structural processing specifically for the action and the audio sequences were identified in bilateral posterior parietal regions (cold colours) and in bilateral temporal regions (hot colours), respectively (lower panel).

Table 1. Congruency effect (incongruent > congruent) in the action imitation and listening tasks.

Region	Hem.	BA	k	x	y	z	Z-value
Action: CONG incongruent > congruent							
Precentral/Inferior Frontal Gyrus	R	44	182	44	6	26	4.29
		44		64	18	24	3.53
		44		54	14	16	3.49
Middle Occipital Gyrus	R	19	352	40	-80	36	4.66
Superior Parietal Lobe	R	7P		32	-78	42	4.32
		7P		16	-70	58	3.54
Superior Parietal Lobe	L	7P	510	-16	-74	58	4.27
		7P		-20	-70	50	3.97
Middle Occipital Gyrus	L	19		-30	-80	34	3.72
Audio: CONG incongruent > congruent							
Inferior Frontal Gyrus	R	45	1667	44	34	2	5.12
		45		44	18	16	4.98
		Insula		36	10	-2	4.49
Superior Temporal Sulcus, pos	R	22	256	48	-32	0	3.92
Superior Temporal Gyrus, pos	R	22		70	-24	6	3.59
		22		60	-34	8	3.46
Cerebellum (Crus II)	L	-	132	-14	-76	-36	4.37

Whole-brain activation cluster sizes (k), MNI coordinates (x, y, z), and Z-scores for the CONG contrast in action imitation and listening tasks ($p_{voxel} < .001$; $p_{cluster} < .05$, FWE corrected). BA: Brodmann area, Hem.: hemisphere.

To identify areas exclusively recruited depending on stimulus format (photos of actions or audio), the CONG contrasts of both tasks were compared using a paired *t*-test (Table 2). Incongruent actions elicited greater activity in bilateral SPL, MOG, and in the left superior/middle frontal gyrus (frontal eye fields, FEF: BA8) (Fig. 2 lower panel, cold colours). Conversely, auditory violations yielded larger BOLD responses in bilateral STS/STG, compared to the action task (Fig. 2 lower panel, hot colours).

To identify areas commonly recruited during both audio and action task, we masked the audio CONG contrast with the action contrast. This analysis yielded a cluster in rIFG (BA44, x = 64, y = 18, z = 24, cluster extent = 28 voxels, Z = 3.53, $p_{voxel} < .001$) that, however, did not survive the cluster-level FWER correction.

Table 2. *t*-test between the CONG contrasts in the action imitation and listening task.

Region	Hem.	BA	k	x	y	z	Z-value
Action > Audio							
Superior Parietal Lobe	R	7	806	26	-76	46	4.46

		7		24	-58	44	4.22
Middle Occipital Gyrus	R	19		34	-80	38	4.08
Middle Occipital Gyrus	L	19	1436	-32	-76	24	4.99
Superior Parietal Lobe	L	7		-30	-74	36	4.75
		7		-20	-72	48	4.41
Superior Frontal Gyrus	L	8	186	-20	2	60	3.97
		8		-22	-2	72	3.60
Audio > Action							
Superior-Middle Temporal Gyrus	R	21/22	1100	64	-16	8	4.79
				62	-32	4	4.64
				52	-30	-8	4.20
Superior-Middle Temporal Gyrus	L	21/22	166	-52	-32	-2	4.11
				-52	-16	2	3.91
				-62	-30	6	3.51

Whole-brain activation cluster sizes (k), MNI coordinates (x, y, z), and Z-scores for the paired-samples *t*-test comparison of the action > audio and audio > action CONG contrast ($p_{voxel} < .001$; $p_{cluster} < .05$, FWE corrected. BA: Brodmann area, Hem.: hemisphere.

3.2.2 Seed-based functional connectivity from the IFG peak maxima.

To gather evidence for the communication between IFG and modality-specific areas, we conducted a seed-based functional connectivity analysis on the resting-state fMRI data acquired from the same pianists. The activation peaks in the IFG clusters in the action imitation and audio task were chosen as seed regions, which were located between the right dorsal BA44 and the pre-central sulcus (action-seed) and in right BA45 (audio-seed). The results are depicted in Figure 3 (upper and middle panels) and show positive functional connectivity (hot colours) between IFG and (amongst others) regions that were functionally specific to the action or audio musical task.

Figure 3

Figure 3. *Upper and middle panels:* seed-based functional connectivity maps on resting-state data from the action seed in dorsal BA44 and audio seed in ventral BA45, respectively. Seeds are depicted as black circles. Hot and cold colours indicate positive and negative functional connectivity, respectively. Consistent with the task-based activations, the topographical connectivity patterns include posterior parietal regions from the action-seed and temporal regions from the audio-seed. *Lower panel: t*-test between the connectivity maps of the action- and audio seed. Correlation values in posterior parietal areas were significantly higher for the action-seed than the audio-seed, whereas temporal regions were more strongly correlated to the audio- than the action-seed.

In line with activity in IFG and posterior parietal regions in the action task, the action-seed in IFG (BA44) exhibited positive correlations within a dorsal motor network comprising bilateral parietal cortex, extending from the anterior ventral supramarginal gyrus (BA40) to the posterior superior parietal lobes (BA7) (Table 3). Notably, there were no significant correlations with temporal regions that were specific to the audio modality. A large cluster peaking in bilateral precentral gyrus showed positive correlations with the action-seed, including subclusters in bilateral BA44 extending to insular regions, ventral premotor cortex (BA6), middle frontal gyrus (BA9) bordering the superior frontal gyrus and the inferior portion of the frontal pole (BA10). Medially, the action seed exhibited positive correlations with the right posterior border of the supplementary motor cortex (BA6) and anterior cingulate (BA24). Finally, there were positive correlations with right inferior temporal gyrus at the temporo-occipital junction (BA20), and bilateral occipitotemporal areas (BA37), cerebellum and thalamus.

Consistent with activity in IFG and temporal areas in the audio task, the audio-seed in IFG (BA45) exhibited positive correlations within the auditory network comprising the posterior part of the right superior temporal gyrus (BA22) and left Heschl's gyrus (including BA41/42) (Table 3). Additionally, there were positive correlations with frontal areas in the right hemisphere including orbitofrontal (BA47/11/12) and frontopolar regions (BA10), superior (BA8), middle frontal areas (BA9), and anterior cingulate gyrus (BA24), and in the left hemisphere, including BA45, BA47, BA9, BA10, and BA12. In the parietal cortex, positive correlations were restricted to bilateral anterior ventral supramarginal gyrus (BA40), without extending to more posterior parietal regions. Finally, there were positive correlations with thalamus and right putamen.

Apart from positive correlations, activity in both action and audio seeds was negatively correlated (Figure 3 upper and middle panels, cold colours) with activity in areas belonging to the default mode network (DMN), namely the cingulate gyrus and the superior portion of bilateral lateral occipital cortex extending into angular gyrus. These regions typically show a decrease of activation during attention-demanding tasks and goal-directed behaviours (Uddin et al., 2009). Additionally, negative correlations were found between the action-seed and bilateral anterior middle temporal gyrus and medial prefrontal cortex, anti-correlations that have been associated with highly difficult goal-directed tasks, as could apply in the case of our action-task (McKiernan et al., 2003).

Finally, a paired-samples *t*-test comparing the connectivity maps of the action- and audio-seed (Table 4) confirmed their differential predominant connectivity to parietal and temporal areas, respectively. Specifically, connectivity of the action-seed (compared to audio-seed) was stronger to bilateral posterior parieto-occipital areas, as well as to bilateral cerebellum, right frontal pole, frontal medial cortex and anterior cingulate gyrus, left superior frontal and precentral gyrus. Conversely, the connectivity of the audio-seed (compared to action-seed) was stronger to bilateral superior or middle temporal gyrus, as well as to bilateral

cerebellum, right superior frontal gyrus, posterior cingulate and angular gyrus, thalamus, and left frontal operculum.

Table 3. Resting-state functional connectivity from the action and audio seed in right inferior frontal gyrus.

		Action-seed					Audio-seed				
Region	BA	k	x	y	z	Z-value	k	x	y	z	Z-value
Right Hemisphere (positive correlations)											
Frontal Pole	10	125	26	38	-16	4.94					
Middle Frontal Gyrus	9						310	42	4	46	5.22
Superior Frontal Gyrus	8						349	4	18	60	4.79
Supplementary Motor Cortex	6	734	6	14	52	5.92					
Cingulate Gyrus, ant.	24	133	4	6	28	5.76	324	4	32	22	4.86
Supramarginal Gyrus/Superior Parietal Lobe	40/7	5418	52	-30	48	6.89					
Superior Temporal Gyrus, post	22						119	50	-14	-8	4.33
Superior Temporal Gyrus, post.	22						1148	52	-30	6	4.72
Middle temporal Gyrus	20/21/37	1304	52	-56	-12	5.92					
Putamen	-						36	32	-12	-8	4.52
Thalamus	-	120	8	-14	8	5.84					
Cerebellum (VIIb)	-	350	18	-68	-48	5.36					
Cerebellum (VI)	-	95	8	-70	-22	4.98					
Left Hemisphere (positive correlations)											
Frontal Pole	10	531	-44	38	8	6.28					
Frontal Operculum Cortex	45						2198	-38	26	0	6.29
Inferior Frontal Gyrus	44						50	-46	12	22	4.29
Precentral Gyrus	6/44/Ins	2415	-44	6	24	6.89					
Middle Frontal Gyrus	9	626	-26	0	50	6.15					
Cingulate Gyrus, ant.	24						52	-2	-12	42	3.98
Heschl's Gyrus (H1 and H2)	41/42						173	-52	-14	4	4.07
Supramarginal Gyrus/Superior Parietal Lobe	40/7	3814	-60	-30	42	5.92	143	-66	-38	26	4.21
Middle temporal Gyrus	20/21/37	1162	-60	-60	-6	5.60	63	-62	-60	8	4.03
Thalamus	-	116	-12	-14	6	4.45	93	-6	-14	2	5.43
Cerebellum (VIIb)	-	969	-26	-66	-52	6.06					
Cerebellum (VI)	-	95	-22	-62	-28	4.98					
Cerebellum (Crus II)	-						58	-16	-78	-34	4.40

Right Hemisphere (negative correlations)											
Frontal Pole	10	11860	6	60	22	7.08					
Superior Frontal Gyrus	8						384	28	30	54	4.97
Middle Temporal Gyrus	21	2367	60	4	-24	6.88					
Cingulate Gyrus, pos.	24	5798	10	-50	34	7.66	3682	10	-46	12	5.35
Cerebellum (IX)	-	187	4	-50	44	5.81					
Cerebellum (Crus I)	-	1799	26	-88	-30	5.99	326	36	-52	-34	4.99
Left Hemisphere (negative correlations)											
Frontal Pole	10						105	-20	64	-6	4.37
Superior Frontal Gyrus	8						601	-20	28	38	5.67
Middle Temporal Gyrus	21	2367	-62	-24	-12	6.88					
Inferior Temporal Gyrus, pos.	20						70	-60	-44	-14	4.17
Lateral Occipital Cortex, sup.	39	1764	-40	-50	26	6.27	1268	-36	-66	38	5.57
Hippocampus	-	62	-34	-34	-8	5.01	65	-30	-34	-12	4.07
Cerebellum (Crus I)	-	27	-44	-56	-42	4.275					

Results of the whole-brain functional connectivity analysis from IFG activation maxima in action imitation and listening tasks. BA: Brodmann area, k: cluster size, MNI coordinates (x, y, z), and Z scores. ($p_{voxel} < .001$; $p_{cluster} < .05$, FWE corrected).

Table 4. Comparison of rs-functional connectivity from the action- and audio-seed in the right IFG.

		Action > Audio seed					Audio > Action seed				
Region	BA	k	x	y	z	Z-value	k	x	y	z	Z-value
Right Hemisphere											
Frontal Pole	10	93	48	42	14	3.56					
Superior Frontal Gyrus	9						4263	4	56	42	5.28
Frontal medial cortex	11	59	4	44	-18	3.88					
Cingulate Gyrus	23/24	52	2	6	30	4.58	200	2	-14	38	4.89
Superior Temporal Gyrus	22						645	52	-8	-8	5.16
Angular Gyrus	40						161	44	-46	32	4.20
Lingual Gyrus	27	72	16	-42	-6	4.87					
Infer. Temporal Gyrus, temp-occ.j.	37	996	52	-56	-14	5.53					
Lateral Occipital Cortex, sup.	7	4448	24	-68	50	6.03					
Precuneus Cortex	17	373	24	-54	18	4.80					
Thalamus	-						50	2	-12	10	4.49
Cerebellum (Crus II)	-	153	4	-78	-44	4.24	68	30	-88	-36	3.86
Cerebellum (XI)	-	51	16	-46	-48	4.46					
Left Hemisphere											
Frontal Operculum Cortex	47						976	-40	26	0	5.35
Superior Frontal Gyrus	8	640	-24	4	52	5.17					
Precentral gyrus	6	549	-52	6	40	5.43					
Middle Temporal Gyrus (middle)	20						206	-56	-20	-12	4.88
Middle Temporal Gyrus (post.)	21						64	-54	-38	0	4.38
Temporal Occipital Fusiform Cortex	7	66	-24	-58	-12	4.10					
Lateral Occipital Cortex, sup.	7	3678	-26	-76	30	5.66					
Lateral Occipital Cortex, inf.	19	567	-50	-76	-4	4.87					
Cerebellum (Crus I/II)	-	114	-6	-76	-40	4.30	168	-24	-76	-34	4.44

Results of the *t*-test between whole-brain functional connectivity from IFG activation maxima in action imitation and listening task. BA: Brodmann area, k: cluster size, MNI coordinates (x, y, z), and Z scores. ($p_{voxel} < .001$; $p_{cluster} < .05$, FWE corrected). temp-occ.j.: temporo-occipital junction.

4. Discussion

The present study investigated the neural bases of action planning and prediction based on long-term knowledge of harmonic regularities and compared them with those involved in auditory prediction. Functional neuroimaging data of expert pianists were acquired at rest, during imitation of (without sound) or listening to (without imitation) harmonically congruent or incongruent chord sequences presented as photos of musical actions or sounds, respectively. Violations in both musical *actions* and *sounds* recruited distinct sub-regions (BA 44 and BA 45, respectively) in right IFG (rIFG) interconnected with parietal visual-motor and temporal auditory areas, respectively. We propose that motoric and auditory long-term representations of harmonic regularities are likely to account for the differential involvement of parietal and temporal areas that enter into dynamic interactions with computations in rIFG. Moreover, the involvement of rIFG in parsing musical action and sound sequences is sensitive to stimulus properties and task – production or perception – accounting for the divergent peak localizations, in line with prevailing models of general prefrontal cortex organization (e.g., Fuster, 2001), and dual stream models of the visuo-spatial (e.g., Goodale & Milner, 1992) and auditory system (e.g., Rauschecker and Scott, 2009). Altogether, our results emphasise dissociable, neural action and audio networks in which modality-specific long-term knowledge and contextual information act as priors for the prediction of forthcoming events. In this respect, predictive coding models (Friston, 2010) may yield a unifying explanatory framework for information processing across both action and perception.

4.1 Musical action

The imitation of incongruent actions elicited activations in fronto-parietal areas (see Table 1), including the right inferior frontal gyrus (IFG: dorsal BA44 extending to the border of the precentral sulcus) and bilateral posterior parietal cortex (pSPL: BA7; MOG: BA19).

This activation pattern resembles the typical dorsal fronto-parietal network for visually guided behaviour that integrates sensory information with action-goals through sensorimotor transformations (Gallivan and Culham, 2015; Kravitz et al., 2011). Accordingly, MOG is known as an area involved in capturing relevant visual-spatial dimensions of objects and visually-guided actions (Lingnau and Downing, 2015). SPL has been associated with high-level aspects of motor behaviour, such as the formation of intentions and early movement plans. These processes are aided by critical operations of multisensory integration and visuomotor transformation in SPL (Andersen and Buneo, 2002). Activations in pSPL have been reported during motor imagery of action-goals and trajectories (Aflalo et al., 2015), attentional spatial remapping/reprogramming of pre-selected actions (O'Reilly et al., 2013), and transformation of spatial target information into corresponding actions (Barany et al., 2014; Schon et al., 2002).

One crucial finding was the recruitment of the rIFG (dorsal BA44) when the final chord, predicted by the harmonic structure of the given musical sequence, was violated. This is consistent with the role of IFG in processing high-level aspects of motor behaviours (Grafton and Hamilton, 2007). Experimental

evidence emphasises the role of bilateral IFG in processing hierarchical relationships within action sequences either when judging complex familiar activities (Farag et al., 2010) or when executing abstract hierarchically organised patterns of action sequences (Koechlin and Jubault, 2006). Altogether, these combined results suggest that the right IFG supports the structural integration of simple acts into more complex combinatorial action sequences. The greater BOLD response during incongruent (compared to congruent) chords may be due to a mismatch with the predicted musical motor act that leads to higher computational costs during structural integration. Importantly, these findings indirectly show that pianists' knowledge of harmonic regularities transfers to the motor domain and enables them to predict and plan forthcoming musical acts during performance.

The absence of auditory activation in the incongruent vs. congruent contrast suggests that pianists relied more on their action knowledge recalled by the execution of the preceding chords than on auditory mechanisms (Bianco et al., 2016; Novembre and Keller, 2011; Sammler et al., 2013). Note that this finding does not conflict with the large body of experimental evidence for action-perception coupling in trained musicians (for review, see Novembre and Keller, 2014; Zatorre et al., 2007) Our unusual and taxing imitation task on unrehearsed sequences may have led pianists to focus on the motor part of the task, possibly suppressing unhelpful auditory images (cf. Pfordresher, 2012; van der Steen et al., 2014)(cf. Pfordresher, 2012). Alternatively, auditory feed-forward mechanisms may not discriminate between congruent and incongruent chords such that auditory activations cancelled out.

Overall, these fronto-parietal activations complement and support our previous behavioural (Novembre & Keller, 2011) and EEG studies on expert pianists (Bianco et al., 2016; Sammler et al., 2013): silent production of harmonically incongruent chords elicited response time costs and a centro-parietal negativity that was associated with mechanisms of motor reprogramming of a pre-planned action in face of the violation. The activations of SPL and MOG match and support our interpretation of the posterior negativity as a correlate of the spatial remapping and reprogramming of pre-planned actions, and the activation of IFG lends evidence that these mechanisms stand under frontal control.

Within the predictive-coding framework (Friston, 2010), a bidirectional flow of information can be suggested to occur in the two hemispheres between parietal areas, processing visual-motor inputs, and the IFG, performing structural integration of incoming items. Indeed, the functional connectivity analysis of our resting state data revealed strong positive correlations between right BA44 and, amongst others, bilateral superior parietal lobes, also revealed by the task-based analysis. A fronto-parietal network relying on the route of the dorsal visual stream has been associated with sensorimotor transformation during visually guided action planning (Goodale and Milner, 1992). According to motor control theory, these operations might be supported by "forward models", through which the expected outcome of an action is compared with actual sensory feedback (Wolpert and Flanagan, 2001). In this framework, posterior parietal regions simultaneously represent potential actions whose pre-selection is biased by the influence of internal models from prefrontal regions (Cisek, 2006). The novel finding is that these internal models may be shaped by the musician's knowledge of harmonic regularities and musical context. We propose

that, on the one hand, visual-motor information about the current act is forwarded from posterior regions to the IFG that integrates the items and builds up an internal model of the sequence's harmonic structure. On the other hand, this internal model affords predictions of visual-spatial surface features of the next chord in MOG and may bias the pre-selection/representation of harmonically appropriate forthcoming motor acts in SPL. The generated model would be continually validated/updated via the matching between the expected action and the combined visual and proprioceptive signals from the current input (Wolpert and Flanagan, 2001). Interestingly, the combined findings raise the hypothesis that (musical) action knowledge, internal visual-motor models and fronto-parietal information flow may provide the basis on which the motor system contributes to visual perception and prediction of human behaviour (Novembre and Keller, 2014).

4.2 Music perception

In line with previous findings (Koelsch et al., 2005), listening to harmonically incongruent compared to congruent chords elicited activations in fronto-temporal areas: right inferior frontal gyrus (IFG: BA44, BA45) extending into the insular cortex, and right posterior superior temporal gyrus and sulcus (pSTG/STS: BA22).

The IFG and the posterior STG have been associated with structural analysis of auditory musical sequences based on internalised knowledge of harmonic regularities (Koelsch et al., 2005; Maess et al., 2001; Sammler et al., 2013a; Tillmann et al., 2006). The IFG has been proposed to support integration of discrete items into higher-order structures, based on which top-down predictions on forthcoming items can be generated. Greater BOLD responses in IFG may reflect the higher computational demand to integrate incongruent chords that are weakly related to the harmonic context and do not fulfil the prediction. Compared to these higher-order computations in IFG, pSTG/STS has been proposed to support lower-level matching processes between the actually perceived and the predicted sensory information (Sammler, Koelsch, et al., 2013). Indeed, posterior superior temporal areas have been associated with physical feature analysis and short-term representation of sounds (Seger et al., 2013), as well as with the identification of the harmonic functions of chords within musical sequences (Musso et al., 2015).

Our connectivity analysis showed a functional coupling between IFG (BA44/BA45) and pSTG/STS, making it plausible to assume bidirectional dynamic fronto-temporal interactions during structural integration processes (Friston, 2010). On the one hand, early sensory analysis of chord functions may be forwarded from temporal to frontal regions where information is structurally integrated and harmonic predictions are established. On the other hand, these predictions may in turn inform the identification process in pSTG/STS where perceived and predicted items are matched to validate or revise the frontal prediction.

Overall, these data emphasise the crucial role of not just one area, but of a dynamic exchange of information between fronto-temporal areas in providing resources for the parsing of complex harmonically organised sounds (Hyde et al., 2011). Neuroanatomically, the fronto-temporal information

exchange may be implemented along dorsal or ventral auditory pathways (see further below) (Loui et al., 2011; Musso et al., 2015; Rauschecker, 2011). The anatomical specification of these pathways, their functional relevance and dependency on musical training are interesting topics for future research.

4.3 Dorsal and ventral streams for musical action and perception

As discussed above, harmonic processing in musical actions and auditory perception relied on dissociable fronto-parietal and fronto-temporal neural networks, respectively. Representations of harmonic regularities in either visual-motor or auditory format are likely to account for the differential involvement of parietal and temporal areas, respectively, that both dynamically interact with computational processes in IFG. Interestingly, these interactions involved distinct posterior-dorsal and anterior-ventral rIFG sub-regions, i.e., BA6/44 in the action imitation task vs. BA44/45 in the audio task. This dissociation may either reflect (i) a task-unspecific sensitivity of IFG to structural processing demands in line with models of general prefrontal cortex specialization, or (ii) a task-specific involvement of dorsal and ventral IFG sub-regions as endpoints of different processing streams.

(i) Investigating harmonic structure processing in perception and action necessarily entails differences in experimental setup that alone suffice to induce different processing demands and shift activation peaks within IFG – even if both peaks may reflect similar structural computations. For example, recent theories propose anterior-posterior (Badre and D'Esposito, 2009; Fuster, 2001; Koechlin and Summerfield, 2007) and/or rostral-caudal (Friederici, 2011) gradients of prefrontal cortex organization along which *similar* functions, e.g., the "integration" of discrete items over time, operate at *different* levels of abstraction (Makuuchi et al., 2012). Along these lines, the more demanding imitation task might have triggered integration over shorter segments in the action sequences (i.e., integration at a lower level of complexity), limiting the activation to dorsal BA44 in the action contrast.

(ii) Alternatively, the divergence of dorsal and ventral rIFG peaks and connectivity profiles may arise from the intrinsically different nature of the tasks – silent musical action imitation vs. listening – in line with dual stream models of the visuo-spatial (Goodale and Milner, 1992; Kravitz et al., 2011) and auditory system (Rauschecker and Scott, 2009; Rauschecker, 2011). According to these models, dorsal portions of IFG are interconnected with the parietal and temporal lobe within dorsal processing streams for time-dependent mechanisms that afford transformation between sensory input (visuo-spatial or sound) and motor representations, thereby supporting action. Dorsal stream involvement has been shown previously for goal-related actions (Kravitz et al., 2011), speech production (Hickok & Poeppel, 2007) and singing (Loui, 2015; Zarate, 2013) and is compatible with our fronto-parietal network observed in pianists during musical action imitation. Moreover, since the audio contrast comprised frontal activation extending to dorsal IFG, it is plausible that also during listening (although without imitation) a dorsal stream of auditory information might have been involved for mapping sound to action simulated by pianists (Zatorre et al., 2007). Notably, the present study adds two new insights: first, we demonstrate that frontal and parietal areas along the dorsal stream provide the neural resources for sequential structure processing

during production of musical sequences; second, unlike in singing or speech production, our action imitation task eliminated auditory feedback during self-produced actions, hence, leading us to conclude that music-structural predictions can be grounded in the visual-motor control system.

Ventral IFG, in turn, is known as endpoint of the auditory ventral stream that, in concert with posterior temporal areas, is classically thought to process pitch information during singing (Berkowska and Dalla Bella, 2009; Zarate, 2013) and to map sound to meaning (Hickok and Poeppel, 2007; Rauschecker and Scott, 2009), compatible with our fronto-temporal network observed during listening. Although musical harmony does not have referential meaning as language, harmonic incongruities do have musical significance to listeners – i.e., intra-musical meaning as framed by Koelsch (2011) – in that the harmonic context leads towards a target chord that can be classified as more or less appropriate for musical closure.

Although the current findings do not speak to the causal role of the nodes or streams, they altogether highlight the relevance of considering structural integration in music production and perception as a network capacity by taking into account the connectivity between frontal computational and posterior modality-specific regions. Flexible and proficient music performance is likely to benefit from the dynamic weighting of these dissociable visual-motor and auditory circuits for prediction and motor planning based on internalised knowledge of harmony.

5. Conclusion

The present data provide first neuroimaging evidence that expert pianists predict forthcoming musical chords not only in auditory perception, but also in the processing of actions independently of auditory information. Remarkably, this suggests that, after intensive training, knowledge of structural regularities influences experts' action planning via implicit mechanisms of motor prediction/control, and might in turn increase proficiency of performance on top of fine movement optimization.

Our paradigm, in which pianists acted without listening to sound and listened without acting, dissociated a dorsal action and a ventral audio network for harmonic prediction, potentially acting in concert during real production (i.e., playing with sound). The dorsal and ventral networks both involve frontal computational sub-regions in rIFG, interconnected with parietal and temporal posterior systems of knowledge, respectively. These networks are likely to provide the infrastructure that allows frontal areas to keep track of abstract dependencies in sequential information via dynamic exchange with progressively lower-level modality-specific systems of knowledge. Predictive coding is proposed as an explanatory framework that unifies both networks' functional roles: to optimise predictions in action and perception based on previous exposure and knowledge of harmony.

Acknowledgments

We are grateful to the pianists who participated in this experiment and to Bruno Pace who lent a hand in stimulus preparation. We also thank Mauricio Martins, Tomás Goucha, Gabriele Chierchia and

616 Emiliano Zaccarella for inspiring discussions and Sven Gutekunst and Jöran Lepsien for technical support.
617 This research was funded by the Max Planck Society.

References

Aflalo, T., Kellis, S., Klaes, C., Lee, B., Shi, Y., Pejsa, K., Shanfield, K., Hayes-jackson, S., Aisen, M., Heck, C., Liu, C., Andersen, R.A., 2015. Decoding motor imagery from the posterior parietal cortex of a tetraplegic human. Science (80-.). 348, 906–910.

Andersen, R.A., Buneo, C.A., 2002. Intentional maps in posterior parietal cortex. Annu. Rev. Neurosci 25, 189–220.

Badre, D., D'Esposito, M., 2009. Is the rostro-caudal axis of the frontal lobe hierarchical? Nat. Rev. Neurosci. 10, 659–669.

Barany, D. a, Della-Maggiore, V., Viswanathan, S., Cieslak, M., Grafton, S.T., 2014. Feature interactions enable decoding of sensorimotor transformations for goal-directed movement. J. Neurosci. 34, 6860–73.

Behzadi, Y., Restom, K., Liau, J., Liu, T.T., 2007. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. Neuroimage 37, 90–101.

Berkowska, M., Dalla Bella, S., 2009. Acquired and congenital disorders of sung performance: A review. Adv. Cogn. Psychol. 5, 69–83.

Bharucha, J., Krumhansl, C.L., 1983. The representation of harmonic structure in music: hierarchies of stability as a function of context. Cognition 13, 63–102.

Bianco, R., Novembre, G., Keller, P.E., Scharf, F., Friederici, A.D., Villringer, A., Sammler, D., 2016. Syntax in Action Has Priority over Movement Selection in Piano Playing: An ERP Study. J. Cogn. Neurosci. 28, 41–54.

Bressler, S.L., Menon, V., 2010. Large-scale brain networks in cognition: emerging methods and principles. Trends Cogn. Sci. 14, 277–290.

Cisek, P., 2006. Integrated neural processes for defining potential actions and deciding between them: a computational model. J. Neurosci. 26, 9761–70.

Eerola, T., Himberg, T., Toiviainen, P., Louhivuori, J., 2006. Perceived complexity of western and African folk melodies by western and African listeners. Psychol. Music 34, 337–371.

Farag, C., Troiani, V., Bonner, M., Powers, C., Avants, B., Gee, J., Grossman, M., 2010. Hierarchical organization of scripts: Converging evidence from fmri and frontotemporal degeneration. Cereb. Cortex 20, 2453–2463.

Fazio, P., Cantagallo, A., Craighero, L., D'Ausilio, A., Roy, A.C., Pozzo, T., Calzolari, F., Granieri, E., Fadiga, L., 2009. Encoding of human action in Broca's area. Brain 132, 1980–8.

Fiebach, C.J., Schubotz, R.I., 2006. Dynamic anticipatory processing of hierarchical sequential events : a common role for broca ' s area and ventral premotor cortex across domains ? Cortex 42, 499–502.

Fitch, W.T., Martins, M.D., 2014. Hierarchical processing in music, language, and action: Lashley revisited. Ann. N. Y. Acad. Sci. 1316, 87–104.

Friederici, A.D., 2011. The brain basis of language processing: from structure to function. Physiol. Rev. 91, 1357–92.

Friston, K., 2010. The free-energy principle: a unified brain theory? Nat. Rev. Neurosci. 11, 127–38.

Fuster, J.M., 2001. The prefrontal cortex - An update: Time is of the essence. Neuron 30, 319–333.

Gallivan, J.P., Culham, J.C., 2015. Neural coding within human brain areas involved in actions. Curr. Opin. Neurobiol. 33, 141–149.

Goodale, M. a., Milner, A.D., 1992. Separate visual pathways for perception and action. Trends Neurosci. 15, 20–25.

Grafton, S.T., Hamilton, A.F.D.C., 2007. Evidence for a distributed hierarchy of action representation in the brain. Hum. Mov. Sci. 26, 590–616.

Hickok, G., Poeppel, D., 2007. The cortical organization of speech processing. Nat. Rev. Neurosci. 8, 393–402.

Hyde, K.L., Zatorre, R.J., Peretz, I., 2011. Functional MRI evidence of an abnormal neural network for pitch processing in congenital amusia. Cereb. Cortex 21, 292–9.

Kilner, J.M., Friston, K.J., Frith, C.D., 2007. Predictive coding: an account of the mirror neuron system. Cogn. Process. 8, 159–66.

Kim, S.-G., Kim, J.S., Chung, C.K., 2011. The effect of conditional probability of chord progression on brain response: an MEG study. PLoS One 6, e17337.

Koechlin, E., Jubault, T., 2006. Broca's area and the hierarchical organization of human behavior. Neuron 50, 963–74.

Koechlin, E., Summerfield, C., 2007. An information theoretical approach to prefrontal executive function. Trends Cogn. Sci. 11, 229–35.

Koelsch, S., 2011. Toward a neural basis of music perception - a review and updated model. Front. Psychol. 2, 110.

Koelsch, S., 2005. Neural substrates of processing syntax and semantics in music. Curr. Opin. Neurobiol. 15, 207–12.

Koelsch, S., Fritz, T., Schulze, K., Alsop, D., Schlaug, G., 2005. Adults and children processing music: an fMRI study. Neuroimage 25, 1068–76.

Koelsch, S., Siebel, W. a, 2005. Towards a neural basis of music perception. Trends Cogn. Sci. 9, 578–84.

Kravitz, D.J., Saleem, K.S., Baker, C.I., Mishkin, M., 2011. A new neural framework for visuospatial processing. Nat. Rev. Neurosci. 12, 217–30.

Krumhansl, C.L., 1983. Perceptual structures for tonal music. Music Percept. 1, 28–62.

Large, E.W., Palmer, C., 2002. Perceiving temporal regularity in music. Cogn. Sci. 26, 1–37.

Lartillot, O., Ayari, M., 2011. Cultural impact in listeners' structural understanding of a Tunisian traditional modal improvisation, studied with the help of computational models. J. Interdiscip. Music Stud. 5, 85–100.

Lerdahl, F., Jackendoff, R., 1983. A generative theory of to- nal music. MA: MIT Press., Cambridge. MA: MIT Press.

Lingnau, A., Downing, P.E., 2015. The lateral occipitotemporal cortex in action. Trends Cogn. Sci. 19,

268–277.

Loui, P., 2015. A Dual-Stream Neuroanatomy of Singing. Music Percept. 32, 232–241.

Loui, P., Li, H.C., Schlaug, G., 2011. White matter integrity in right hemisphere predicts pitch-related grammar learning. Neuroimage 55, 500–507.

Loui, P., Wu, E.H., Wessel, D.L., Knight, R.T., 2009. A generalized mechanism for perception of pitch patterns. J. Neurosci. 29, 454–459.

Maess, B., Koelsch, S., Gunter, T.C., Friederici, A.D., 2001. Musical syntax is processed in Broca ' s area : an MEG study. Nat. Neurosci. 4, 540–545.

Maidhof, C., Vavatzanidis, N., Prinz, W., Rieger, M., Koelsch, S., 2009. Processing Expectancy Violations during Music Performance and Perception : An ERP Study. J. Cogn. Neurosci. 22, 2401–2413.

Makuuchi, M., Bahlmann, J., Friederici, A.D., 2012. An approach to separating the levels of hierarchical structure building in language and mathematics. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 367, 2033–2045.

Martin, J. a., Karnath, H.-O., Himmelbach, M., 2015. Revisiting the cortical system for peripheral reaching at the parieto-occipital junction. Cortex 64, 363–379.

McKiernan, K. a, Kaufman, J.N., Kucera-Thompson, J., Binder, J.R., 2003. A parametric manipulation of factors affecting task-induced deactivation in functional neuroimaging. J. Cogn. Neurosci. 15, 394–408.

Musso, M., Weiller, C., Horn, A., Glauche, V., Umarova, R., Hennig, J., Schneider, A., Rijntjes, M., 2015. A Single Dual-Stream Framework for Syntactic Computations in Music and Language. Neuroimage 117, 267–283.

Novembre, G., Keller, P.E., 2014. A conceptual review on action-perception coupling in the musicians' brain: what is it good for? Front. Hum. Neurosci. 8, 603.

Novembre, G., Keller, P.E., 2011. A grammar of action generates predictions in skilled musicians. Conscious. Cogn. 20, 1232–43.

O'Reilly, J.X., Schüffelgen, U., Cuell, S.F., Behrens, T.E.J., Mars, R.B., Rushworth, M.F.S., 2013. Dissociable effects of surprise and model update in parietal and anterior cingulate cortex. Proc. Natl. Acad. Sci. U. S. A. 110, E3660–9.

Palmer, C., van de Sande, C., 1995. Range of planning in music performance. J. Exp. Psychol. Hum. Percept. Perform. 21, 947–62.

Patel, A.D., 2003. Language, music, syntax and the brain. Nat. Neurosci. 6, 674–81.

Pfordresher, P.Q., 2012. Musical training and the role of auditory feedback during performance. Ann. N. Y. Acad. Sci. 1252, 171–178.

Power, J.D., Schlaggar, B.L., Petersen, S.E., Jonathan D. Power, Bradley L. Schlaggar, Steven E. Petersen, 2015. Recent progress and outstanding issues in motion correction in resting state fMRI. Neuroimage 105, 536–551.

Rauschecker, J.P., 2011. An expanded role for the dorsal auditory pathway in sensorimotor control and integration. Hear. Res. 271, 16–25.

Rauschecker, J.P., Scott, S.K., 2009. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat. Neurosci. 12, 718–24.

Rohrmeier, M. a, Koelsch, S., 2012. Predictive information processing in music cognition. A critical review. Int. J. Psychophysiol. 83, 164–75.

Rohrmeier, M., Rebuschat, P., 2012. Implicit Learning and Acquisition of Music. Top. Cogn. Sci. 4, 525–553.

Ruiz, M.H., Jabusch, H.-C., Altenmüller, E., 2009. Detecting wrong notes in advance: neuronal correlates of error monitoring in pianists. Cereb. Cortex 19, 2625–39.

Sammler, D., Koelsch, S., Ball, T., Brandt, A., Grigutsch, M., Huppertz, H.-J., Knösche, T.R., Wellmer, J., Widman, G., Elger, C.E., Friederici, A.D., Schulze-Bonhage, A., 2013a. Co-localizing linguistic and musical syntax with intracranial EEG. Neuroimage 64, 134–46.

Sammler, D., Koelsch, S., Friederici, A.D., 2011. Are left fronto-temporal brain areas a prerequisite for normal music-syntactic processing? Cortex. 47, 659–73.

Sammler, D., Novembre, G., Koelsch, S., Keller, P.E., 2013b. Syntax in a pianist's hand: ERP signatures of 'embodied' syntax processing in music. Cortex. 49, 1325–39.

Schon, D., Anton, J.L., Roth, M., Besson, M., 2002. An fMRI study of music sight-reading. Neuroreport 13, 2285–2289.

Seger, C.A., Spiering, B.J., Quraini, A.G.S.S.I., Alpeter, C., David, J., Thaut, M.H., Abstract, 2013. Corticostriatal Contributions to Musical Expectancy Perception. J. Cogn. Neurosci. 25:7, 1062–1077.

Smith, S.M., Fox, P.T., Miller, K.L., Glahn, D.C., Fox, P.M., Mackay, C.E., Filippini, N., Watkins, K.E., Toro, R., Laird, A.R., Beckmann, C.F., 2009. Correspondence of the brain ' s functional architecture during activation and rest. PNAS 106, 13040–13045.

Swain, J.P., 1995. The concept of musical syntax. Music. Q. 79, 281–308.

Tenenbaum, J.B., Kemp, C., Griffiths, T.L., Goodman, N.D., 2011. How to grow a mind: statistics, structure, and abstraction. Science 331, 1279–85.

Tillmann, B., 2012. Music and Language Perception: Expectations, Structural Integration, and Cognitive Sequencing. Top. Cogn. Sci. 4, 568–584.

Tillmann, B., Bharucha, J.J., Bigand, E., 2000. Implicit learning of tonality: {A} self-organizing approach. Psychol. Rev. 107, 885–913.

Tillmann, B., Koelsch, S., Escoffier, N., Bigand, E., Lalitte, P., Friederici, a D., von Cramon, D.Y., 2006. Cognitive priming in sung and instrumental music: activation of inferior frontal cortex. Neuroimage 31, 1771–82.

Tymoczko, D., 2003. Function theories: A statistical approach. Musurgia 10, 35–64.

Uddin, L.Q., Kelly, A.M.C., Biswal, B.B., Castellanos, F.X., Milham, M.P., 2009. Functional

Connectivity of Default Mode Network Components: Correlation, Anticorrelation, and Causality. Hum. Brain Mapp. 30, 625–637.

van der Steen, M.C., Molendijk, E.B.D., Altenmueller, E., Furuya, S., 2014. Expert pianists do not listen: The expertise-dependent influence of temporal perturbation on the production of sequential movements. Neuroscience 269, 290–298.

Vuust, P., Witek, M. a. G., 2014. Rhythmic complexity and predictive coding: a novel approach to modeling rhythm and meter perception in music. Front. Psychol. 5, 1–14.

Wolpert, D.M., Flanagan, J.R., 2001. Motor prediction. Curr. Biol. 11, R729–R732.

Zarate, J.M., 2013. The neural control of singing. Front. Hum. Neurosci. 7, 237.

Zatorre, R.J., Chen, J.L., Penhune, V.B., 2007. When the brain plays music: auditory-motor interactions in music perception and production. Nat. Rev. Neurosci. 8, 547–58.

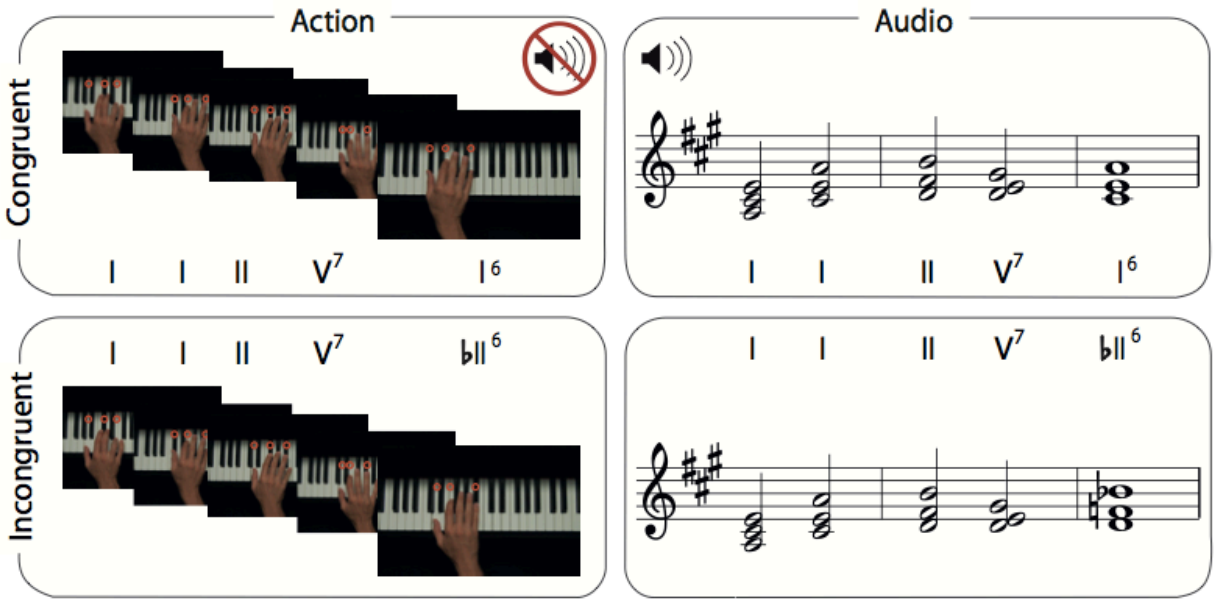

Figure 1. Experimental design: expert pianists were presented with harmonic congruent or incongruent chord progressions, presented either as muted musical actions (photos of a hand playing chords) that they were required to imitate on a glass-board (left panel), or in an auditory format that they listened to (right panel).

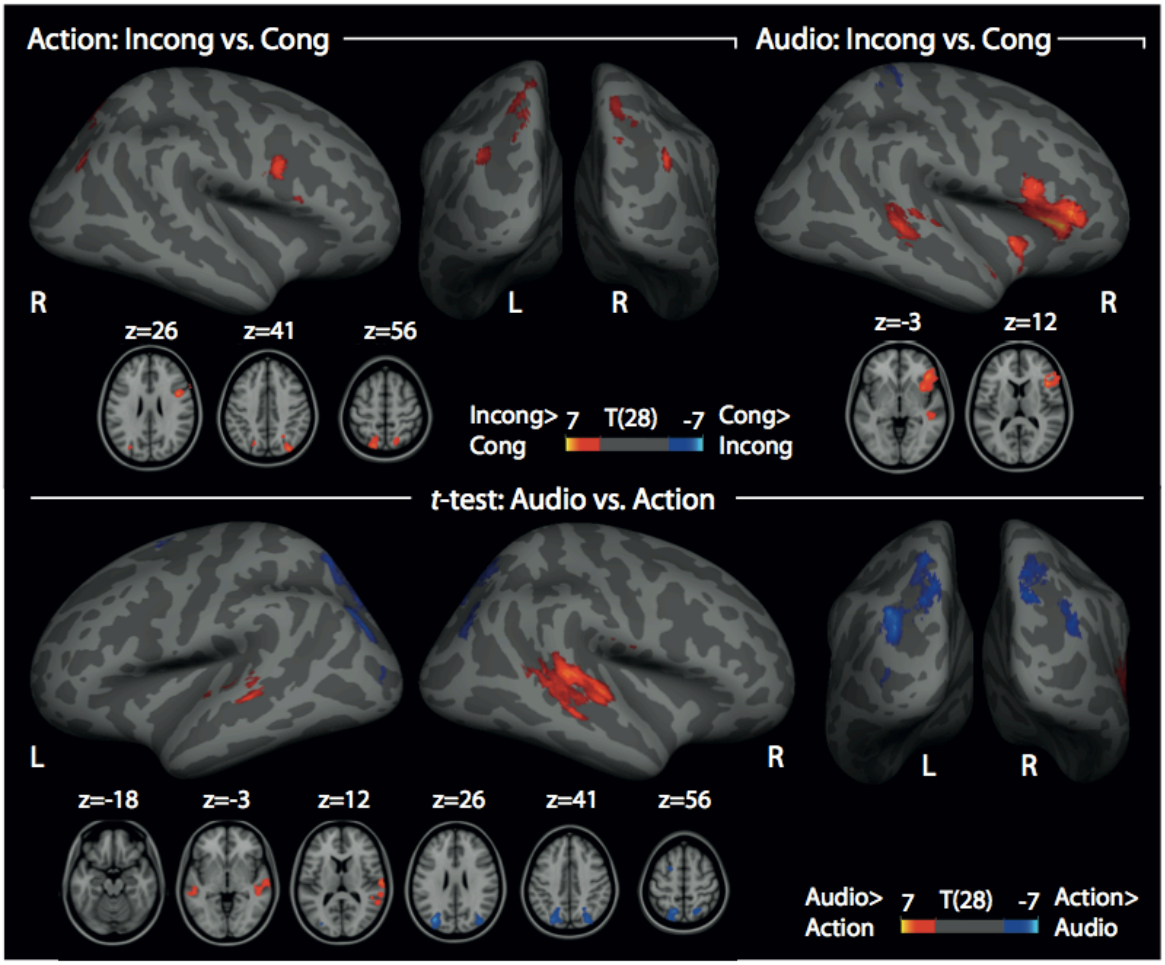

Figure 2. Harmonic violations elicited activations in fronto-parietal areas during action imitation (upper left panel) and in fronto-temporal areas during listening (upper right panel). Areas involved in structural processing specifically for the action and the audio sequences were identified in bilateral posterior parietal regions (cold colours) and in bilateral temporal regions (hot colours), respectively (lower panel).

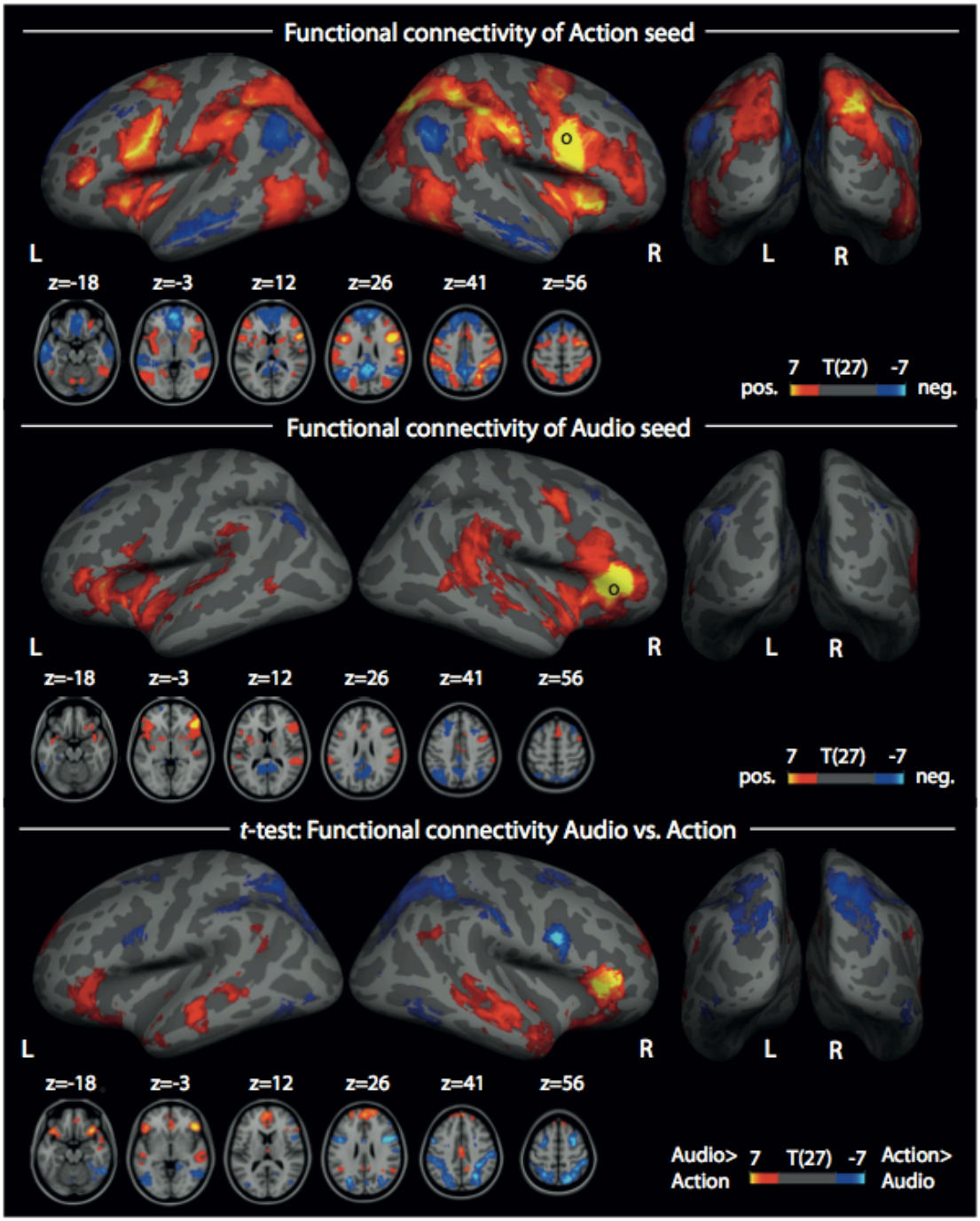

Figure 3. *Upper and middle panels:* seed-based functional connectivity maps of resting-state data from the action seed in dorsal BA44 and audio seed in BA45, respectively. Seeds are depicted as black circles. Hot and cold colours indicate positive and negative functional connectivity, respectively. Consistent with the task-based activations, the topographical connectivity patterns include posterior parietal regions from the action-seed and temporal regions from the audio-seed. *Lower panel: t*-test between the connectivity maps of the action- and audio seed. Correlation values in posterior parietal areas were significantly higher for the action-seed than the audio-seed, whereas temporal regions were more strongly correlated to the audio- than the action-seed.