The importance of integration and top-down salience when listening to complex multi-part musical stimuli

Marie Uhlig [a,*], Merle T. Fairhurst [a,b], Peter E. Keller [a,c]

[a] Max Planck Institute for Human Cognitive and Brain Sciences, Research Group, Music Cognition and Action, D-04103 Leipzig, Germany
[b] Max Planck Institute for Human Cognitive and Brain Sciences, Research Group, Early Social Development, D-04103 Leipzig, Germany
[c] The MARCS Institute, University of Western Sydney, Sydney, Australia

ARTICLE INFO

ABSTRACT

In listening to multi-part music, auditory streams can be attended to either selectively or globally. More specifically, musicians rely on prioritized integrative attention which incorporates both stream segregation and integration to assess the relationship between concurrent parts. In this fMRI study, we used a piano duet to investigate which factors of a leader–follower relationship between parts grab the listener's attention and influence the perception of multi-part music. The factors considered included the structural relationship between melody and accompaniment as well as the temporal relationship (asynchronies) between parts. The structural relationship was manipulated by cueing subjects to the part of the duet that had to be prioritized. The temporal relationship was investigated by synthetically shifting the onset times of melody and accompaniment to either a consistent melody or accompaniment lead. The relative importance of these relationship factors for segregation and integration as attentional mechanisms was of interest. Participants were required to listen to the cued part and then globally assess if the prioritized stream was leading or following compared to the second stream. Results show that the melody is judged as more leading when it is globally temporally ahead whereas the accompaniment is not judged as leading when it is ahead. This bias may be a result of the interaction of salience of both leader–follower relationship factors. Interestingly, the corresponding interaction effect in the fMRI-data yields an inverse bias for melody in a fronto-parietal attention network. Corresponding parameter estimates within the dlPFC and right IPS show higher neural activity for attending to melody when listening to a performance without a temporal leader, pointing to an interaction of salience of both factors in listening to music. Both frontal and parietal activation implicate segregation and integration mechanisms and a top-down influence of salience on attention and the perception of leader–follower relations in music.

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

Our auditory environment consists of complex scenes that have to be analyzed in parts or as a whole. Multi-part music is an example of a complex auditory scene that can either involve focusing on a particular stream or listening holistically to all parts. Mechanisms such as auditory stream segregation allow the brain to separate different sound sources and make it possible to selectively attend to them individually. In order to make sense of a complete auditory scene however, it is necessary also to compare or integrate its composite parts (Nelken, 2011). In the following functional magnetic resonance imaging (fMRI) study, we explore the neural underpinnings of these two attentional mechanisms and how they are differentially employed when listening to and assessing a piano duet with respect to leader–follower relations.

Auditory stream segregation and integration are equally important in the context of musical ensemble performance in which players have to simultaneously attend to different auditory streams including their part (Bigand et al., 2000; Keller, 2001, 2008). Keller (2008) hypothesized that musicians need to employ a specialized form of prioritized integrative attention in order to achieve high synchronization within an ensemble. Attentional resources would be divided between the prioritization of one's own playing and the simultaneous integration of co-performers' sounds in order to match and adjust one's playing for synchronization. In addition, Bigand et al. (2000) were able to show that musicians tend to integrate two parts of multi-part music rather than to divide their attention between them. In an error detection task in which two unknown melodies were concurrently played, musicians' false alarms suggested this kind of listening strategy for multi-part music (Bigand et al., 2000).

The integration of different musical streams thus relies on specific attentional and perceptual processes and is necessary both for synchronized group music making as well as listening to multi-part music. An integration process combines the auditory streams in a

* Corresponding author at: Max Planck Institute for Human Cognitive and Brain Sciences, Research Group, Music Cognition and Action, Stephanstrasse 1a, D-04103 Leipzig, Germany.
E-mail address: uhlig@cbs.mpg.de (M. Uhlig).

common representational space for perception of a global sound. A coherent soundscape nevertheless not only includes the combination of streams but also an assessment of their relationship to each other (i.e. temporally, harmonically, etc.) (Bigand et al., 2000; Bregman, 1990; Erickson, 1975).

How complex musical streams are processed during auditory stream analysis will thus depend on the nature of the relationship between the component parts in a musical piece. The structural relationship of the music may be one factor that influences how the streams are individually or globally perceived and assessed. Structurally speaking, in much western music the melody generally dominates over the accompanying harmony (Bregman, 1990; Erickson, 1975). This hierarchal relationship describes the melody as a structurally independent stream while the accompaniment plays a supporting role for the completion or complementation of the melody. These roles characterizing melody and accompaniment have been described as analogous to figure-ground perception (Tagg, 2003a,b), where the melody is the figure and the accompaniment serves as its background. Such hierarchical structuring may of course vary in its degree, but is nevertheless a defining feature of the music's compositional structure and perceptual organization (Erickson, 1975; Tagg, 2003a,b). The prioritization of the melody might additionally be influenced by perceptual salience factors such as pitch height or more complex rhythms (McAdams and Drake, 2002). Within this structural relationship, it is thus generally the case that the melody can be described as 'globally leading' (to the extent that it dominates perception) and the accompaniment as 'following'.

Another factor that can affect the way in which multi-part music is perceived and assessed in terms of leader–follower relations is the temporal relationship between parts i.e. the accuracy with which the notes of different parts are played together. Simply put, one part – for intentional or unintentional reasons – may be played temporally ahead or behind that of others and would, as such, be heard as temporally leading or lagging, respectively (Goebl and Palmer, 2009; Palmer, 1997; Rasch, 2000; Repp, 1996). Which part is intended to lead temporally is a matter of musical style or interpretation (Rasch, 2000; Repp, 1996). Listeners of western classical music might thus be more familiar with a melody lead, whereas jazz fans might be more accustomed to an accompaniment lead. Unintentional timing errors can also result in one player being ahead of the other. Temporal-leader follower relations are beneficial (regardless of whether they are intentional or unintentional), as it has been shown that a certain degree of asynchrony between parts facilitates the perception of separate tones and is required for stream segregation (Handel, 1989; Rasch, 1979; Wright and Bregman, 1987).

Both the structural and the temporal relationship influence the perceived association between parts and can describe a leader–follower relationship in music. Which of these two factors capture our attention when listening to multi-part music has not yet been investigated.

Based on recent studies, we have an understanding of the neural underpinnings involved in selective attention to multi-part music (Janata et al., 2002; Satoh et al., 2001). However little is known about the relative importance of the integration and perception of different types of leader–follower relationships between parts. Importantly, the tasks used in former studies either involved a target detection task or instructions to selectively listen to one part while ignoring the rest. Considering the importance of the integration of parts, not only when playing but also when listening to multi-part music, a task which allows for prioritized integrative attention mechanisms seems to better capture processes involved in music listening (Bigand et al., 2000; Keller and Burnham, 2005; Nelken, 2011). Such a task enables the listener to prioritize one part while still integrating the other part(s) into a coherent soundscape. This more naturalistic way of listening to music facilitates perception of relationships between parts, which is an important component of multi-part music (Bigand et al., 2000; Erickson, 1975). Moreover, although useful for exploring a factor of selective attention, some of

the musical stimuli used in earlier studies were synthetically generated and thus had no asynchrony between the different instrumental parts (Janata et al., 2002). In the present study, we therefore more specifically explore the neural correlates of attentional mechanisms used when listening to excerpts from an original performance and from corresponding manipulated stimuli derived from this performance (Janata et al., 2002).

To do so, we implemented a cued attention task allowing us to examine the prioritized integrative attention process involved in listening to multi-part music (Bigand et al., 2000; Keller, 2001, 2008). After listening to a recording of a piano duet with a clear structural relationship (melody vs. accompaniment), subjects were asked to globally assess the relative leader–follower relationship of two parts which made up the stimulus as well as its performance quality and the difficulty of the task. The global relationship assessment necessitated subjects not to attend selectively to the cued part but rather to prioritize it and additionally to integrate the second part. We also included stimuli in which we had shifted onset times of either the melody or the accompaniment part by a fixed amount so that one part was consistently temporally leading. This manipulation thus allowed us to look not only at the influence of a structural but also of the temporal relationship between parts on overall perception of a leader–follower relationship. Specifically, due to the combination of the prioritized integrative attention task and the global assessment of the relationship between parts, we were able to investigate how integration and segregation differ in terms of their neural representation.

As our task required the segregation, organization and integration of diverse aspects of auditory information, we hypothesized the recruitment of the intraparietal sulcus (IPS). Its role in organizing sensory information makes it a prime candidate for the organization of top-down and bottom up information for stream integration (Alexander et al., 2005; Champod and Petrides, 2007; Cusack, 2005; Donner et al., 2002; Foster and Zatorre, 2010; Hill and Miller, 2010; Shafritz et al., 2002; Wei et al., 2011; Zatorre et al., 2010). Stream segregation was expected to mostly involve activation of frontal areas typically seen during working memory tasks as well as in instances of sustained attention (Gaab et al., 2003; Pallesen et al., 2010; Strait and Kraus, 2011). However, as our attention task necessitated subjects to segregate as well as integrate concurrent streams, we expected an interaction of both listening strategies on a neural level. Moreover, a top-down influence for both listening styles via a fronto-parietal attention network was expected (Champod and Petrides, 2007; Corbetta and Shulman, 2002).

Both relationship factors seem important to the production and perception of multi-part music (Bregman, 1990; Goebl and Palmer, 2009; Handel, 1989; Rasch, 1979; Wright and Bregman, 1987). We therefore predicted that both factors would influence attention and thereby the perception and assessment of the relationship between parts. Nevertheless, the individual salience of these factors could still differ. As the stimuli consisted of a western style classical duet, familiarity with melody lead might bias perception and underlying neural correlates. It was also possible that the salience of a part of the duet might interact with the attention task of this study. As both factors may drive attention when listening to music, an interaction of both factors and thus an interaction of their salience was expected to shape the subjective leader–follower rating of the perceived music and maybe even the underlying neural activity (Reddy et al., 2009; Reynolds and Desimone, 2003). Top-down modulatory effects related to increases in salience have been shown to involve a fronto-parietal network, including the dorsolateral prefrontal cortex (dlPFC) and the IPS (Bressler et al., 2008; Corbetta and Shulman, 2002). Such a difference in salience of the two factors might furthermore lead to interference and consequently greater difficulty in the attention task (Lavie and De Fockert, 2005; Lavie et al., 2004). We thus additionally expected a salience difference of the two relationship factors to increase cognitive load and influence BOLD activation (Adler et al., 2001; Pugh et al., 1996). Acquired difficulty ratings were used to disentangle effects of salience and cognitive load.

2. Materials and methods

2.1. Subjects

Seventeen (8 female) right-handed healthy volunteers with a mean age of 26.12 years (SD ± 4.2) were recruited for this study. All except one subject were experienced pianists with average of 16.44 (SD ± 5.92) years of playing experience. The exceptional subject was a musician with 10 years of clarinet and guitar experience. Subjects signed a written consent form as part of Max–Planck Institute protocol and were paid for their participation. All had participated in a behavioral pilot study and were thus familiar with the musical stimuli. This ensured that they were capable of distinguishing between the two parts (melody and accompaniment) of the duet stimuli.

2.2. Design & procedure

The design (Fig. 1A) of the study was organized into stimulus and rating phase. Subjects were instructed and cued to listen to one part of a piano duet. To manipulate prioritized integrative attention, the intensity of the not to be attended part was faded in over five seconds thus cueing the participants to the prioritized stream. Participants then continued to attend to the selected part for a further 20 s with all stimuli lasting a total of 25 s (Fig. 1A). After listening to the stimulus, subjects were asked to provide ratings for two out of three possible judgments. Ultimately, they had to assess (1) the leader–follower

relationship of the part they had just attended to relative to the other part, (2) the overall experienced performance quality (this was not intended to serve as a question for an emotional judgment but rather as a rational esthetic and expertise judgment) and (3) an assessment of the individual difficulty of the task during the preceding stimulus. Both the leader–follower rating and the assessment of quality required integrative appraisal of the two parts. Additionally, it is important to note that we did not mention the two possible factors of the leader–follower relationship to avoid influencing their salience and hence the behavioral and neural responses to them.

All global ratings were given on a visual analog scale within an eight second time window. Data were then converted to an 11 point Likert scale. Scales were labeled "*relationship*" ("Verhältnis") with the two anchors "leading" ("anführend") and "following" ("folgend"), "*difficulty*" ("Schwierigkeit") with the anchors "easy" ("leicht") and "very hard" ("sehr schwer") and "*quality*" ("Qualität") with the two anchors "good" ("gut") and "poor" ("schlecht"). The selection and order of the presented ratings were randomly assigned. Subjects practiced giving responses during a short pre-scan trial after lying down on the scanner bed. A two-button response device was used to move the cursor along the scale with either single or continuous presses by the index and middle finger.

Each stimulus was preceded by a white fixation cross in the center of the screen during which participants were instructed not to react. During stimulus presentation the fixation cross changed to green. The stimulus presentation was then followed by a white fixation cross for 11 s
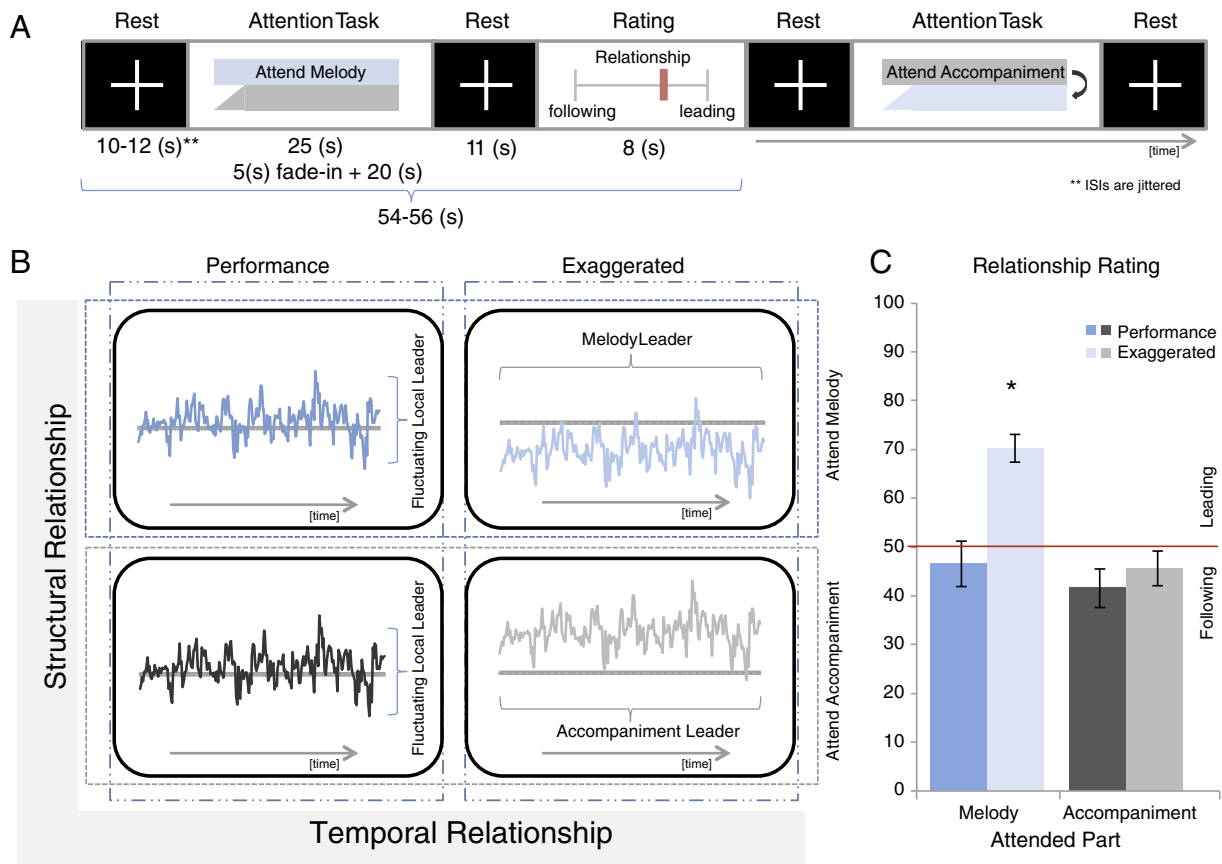

Fig. 1. Study design and behavioral results. (A) Experimental paradigm. In each trial, subjects were cued to attend to one part of the duet while the second part of the auditory stimulus was gradually faded-in over five seconds. The duet then lasted a further 20 s. Each stimulus was followed by a rating period. (B) The 2 × 2 factorial design with the factors "structural relationship" and "temporal relationship" was comprised of the conditions: attend melody in performance stimulus (MP), attend melody in exaggerated stimulus (ME), attend accompaniment in performance stimulus (AP) and attend accompaniment in exaggerated stimulus (AE). Diagrams show a graphical representation of the temporal relationship (asynchronies) between parts in performance stimuli (left) or the shifted (28 ms) exaggerated stimuli (right) in either the attend to melody (top) or attend to accompaniment (bottom) conditions. (C) Group mean leader–follower relationship ratings (and standard errors) for MP (blue), ME (light blue), AP (gray), and AE (light gray). Values >50 indicate the attended part to be subjectively leading, and ratings <50 indicate it to be following.

before the first of the two Likert-scales appeared, cued by its heading ("relationship", "difficulty", or "quality"). The experiment included nine repeats for each stimulus and lasted about 55 min. It was controlled using Presentation software from Neurobehavioral Systems (http://www.neurobs.com/).

In a questionnaire administered in post-scan interviews none of the participants reported having trouble focusing on the to-be-attended part, or being distracted by the fade-in in the questionnaire.

2.3. Stimuli

All stimuli were derived from the same excerpt of a short piano duet, entitled 'the Sicilian hunting song', by Ottorino Respighi. The duet consists of two equal halves as it is repeated with only a different cadence at the end. The original performance, produced by two conservatorium-level pianists, of the entire duet lasted 50 s. The 25-second long excerpt used for the stimuli was the repetition or second half of the duet in order to end the stimulus with a perfect cadence (giving a sense of closure).

This song presents a clear hierarchical structure, where the melody resides within one part of the duet, in a higher frequency range, and the accompaniment within the second part during the entire excerpt, making the melody the global structural leader and the accompaniment the global structural follower (Bregman, 1990; Erickson, 1975). A human performance of the piece was recorded and corrected for a small number of wrong or missing tones. This so-called "original" performance stimulus however included all temporal and velocity (i.e. force) variations natural to a human performance. Globally, the recording had no temporal leader (i.e., the median asynchrony was zero), presumably due to the musical style and the performers' interpretive preferences. However, the local inter-part temporal fluctuations or asynchronies lead to one or the other part being temporally ahead in their onset times. Thus the two parts locally alternated temporal leader follower roles (Fig. 1B).

The two-factorial design was set by the factors of structural relationship and temporal relationship between parts, where the former was manipulated by the task and the latter by the stimulus type. Combining the performance stimulus (P) with the attention task results in two conditions: attend melody (M) or attend accompaniment (A) when listening to the performance stimulus (MP & AP) (Fig. 1B). To contrast this, we included two conditions in which to artificially create a global leader from this recording, we shifted all note onsets of one part 28 ms to be temporally ahead relative to the other part. This exaggeration (E) made either the melody (-28 ms) globally leading or the accompaniment ($+28$ ms) globally leading. In these two conditions the attended stream was always temporally exaggerated (ME & AE). The degree of this temporal lag (i.e. asynchrony) is well within the range of reported related studies (Goebl and Palmer, 2009; Keller and Appel, 2010). Specifically, 28 ms are within the range of natural performance asynchronies, taking into account a combination of different performance situations like playing different pieces with and without visual contact (Keller and Appel, 2010).

By shifting each onset time by the same amount, the local fluctuations were preserved and just the amount of variance was locally exaggerated. Using the same recording for both kinds of stimuli ensured that no other factors would be a potential influence as well as that no cue would differentiate the stimuli apart from our manipulation. For both parts, the same timbre was used to avoid possible confounds of timbre and stimulus complexity. The stimuli were recorded using maxMSP and stored as Musical Instrument Digital Interface (MIDI) files. Files were then converted to .csv text files using an in-house Perl script. After editing the respective onset times, files were converted back to MIDI before creating wave files with Finale® (MakeMusic Inc., USA) using its *Grand Piano* timbre. We additionally included a high baseline control stimulus which consisted of a metronomic computer generated version of the song. Results of this control stimulus will be reported elsewhere.

Stimuli were presented at a constant average intensity of 80 dB with an audio system customized for use in high magnetic fields (MR-Confon http://www.mr-confon.de/en/). Earplugs combined with the audio system's muffled headphones passively attenuated the scanner noise. Subjects were familiarized with both the task and the stimuli prior to scanning to ensure they were able to fulfill task requirements.

2.4. Imaging

Magnetic resonance images were acquired using a Siemens 3-T Tim trio scanner with a standard bird cage head coil. Foam pillows and paddings were used to minimize head motion. Images were obtained continuously during functional scanning using a gradient-echo, echo-planar pulse sequence (TR = 2 s; TE = 28 ms; 31 coronal oblique slices with a one millimeter gap; $3 \times 3 \times 3$ mm in-plane resolution). Imaging parameters were selected to minimize rhythmic noise bursts from the scanner and therefore a possible influence on task difficulty. In a short behavioral pilot experiment, a number of participants went through the task with and without a recording of the scanner noise presented in a regular laboratory setting outside the scanner without any effects on difficulty, thus controlling for the effect of scanner noise. To allow for longitudinal magnetization to approach equilibrium the first four volumes of each functional run were discarded.

2.5. Data analysis

2.5.1. Behavioral data

We calculated a group mean (across subjects) of the four conditions for the corresponding rating-responses. Repeated measures analyses of variance (ANOVAs) were run on each of the three ratings, verifying the directions of the significant effects through separate post-hoc t-tests (corrected for multiple comparisons with an $\alpha = 0.025$) using SPSS.

2.5.2. Imaging data

Analysis of all neuroimaging data sets was performed using FEAT (FMRIB Expert Analysis Tool) Version 5.98, part of FSL (FMRIB's Software Library, www.fmrib.ox.ac.uk/fsl). Pre-statistic processing included: motion correction using MCFLIRT (Motion Correction FMRIB's Linear Image Registration tool), (Jenkinson and Smith, 2001), non-brain removal using BET (Smith, 2002), spatial smoothing using a Gaussian Kernel of 5 mm full width at half-maximum and non-linear high pass temporal filtering (Gaussian-weighted least-squares straight line fitting, with sigma = 40.0 s). Registration included co-registration of the functional scan onto the individual T1 high-resolution structural image and then registration onto a standard brain (Montreal Neurological Institute MNI 152 brain) using FLIRT (FMRIB's Linear Image Registration Tool (Jenkinson and Smith, 2001)). Statistical analysis at the individual subject level was carried out using a general linear modeling (GLM) approach (Friston et al., 1994). Time-series statistical analysis was carried out using FILM (FMRIB's Improved Linear Model) with local autocorrelation correction (Woolrich et al., 2001). This analysis method allows for incorporation of variance within session and across time (fixed effects) and cross session variances (random effects). Cluster thresholding was performed with a Z-threshold of 2.3 and a corrected p-value of <0.01 with a cluster-based correction for multiple comparisons using Gaussian Random Field Theory (Friston et al., 1994; Worsley et al., 1992). In a second step, the difficulty rating was used to check imaging results for an effect of difficulty or cognitive load, and was entered as a regressor into the GLM analysis. (Results for this second analysis will be mentioned in the results section, but only fully shown in the supplementary material.)

2.5.2.1. ANOVA and ROI analysis.
We used a 2×2 (structural relationship × temporal relationship) ANOVA model in a second level fixed effects analysis by constructing an F-contrast. We tested for main effects of both factors as well as a structural relationship

by temporal relationship interaction, which would then indicate an overall leader–follower relationship network. Extracted mean individual parameter estimates (within a sphere with an 8 mm radius around the peak voxel) were used to demonstrate the direction of the interaction. Individual parameter estimates were extracted from regions of interest using PEATE — Perl Event-related Average Time course Extraction (see http://www.jonaskaplan.com/peate/).

3. Results

3.1. Behavioral results

The 2 × 2 ("structural relationship" × "temporal relationship") repeated measures ANOVA on our subjective leader–follower relationship ratings showed a main effect of structural musical relationship, i.e. attending to either the melody or the accompaniment ($F(1,16) = 14.97$, $p = 0.001$) and a main effect of temporal manipulation of the leader–follower relationship between parts ($F(1,16) = 13.388$, $p = 0.002$). The observed interaction between our attentional manipulation and the temporal relationship ($F(1,16) = 8.31$, $p = 0.011$) is driven by a bias to rate the melody part as more leading when it is temporally globally ahead (ME: ($t(16) = 7.15$, $p = 0.000$)). This is despite the fact that subjects did not rate the exaggerated stimulus with the global accompaniment lead (AE) as leading ($t(16) = 1.23$, $p = 0.235$, n.s.) (Fig. 1C). As structural relationship and temporal relationship are factors between the parts and not within a part, the results are indicative of an effect of integration rather than segregation. Post-hoc t-tests showed a significant difference between the two exaggerated stimuli (ME & AE $t(16) = 4.79$, $p = 0.000$) and between the two temporal manipulations when attending to melody (MP & ME: $t(16) = 3.72$, $p = 0.002$).

The same bias becomes obvious when comparing the quality ratings for attending to melody and accompaniment. While the effect for structural relationship in the 2 × 2 ANOVA only approached significance ($F(1,16) = 4.01$, $p = 0.062$), post-hoc paired t-tests (2-tailed) comparing the quality ratings between the different conditions show that the effect is driven by the exaggerated melody lead (ME) to have greater perceived quality (ME & AE: $t(16) = 3.44$, $p = 0.003$; MP & ME: $t(16) = 2.63$, $p = 0.018$) (please see supplementary material, Fig. S1D). Here again the subjective rating of perceived quality necessitates online integration of both the prioritized attended stream and the second stream in order to perceive a global sound or coherent soundscape. Thus the structural bias we see which is boosted through our temporal manipulation may be due to the integration process.

To control for differences in perceived task difficulty we additionally acquired subjective ratings for each condition. A significant interaction ($F(1,16) = 14.36$, $p = 0.002$) of structural relationship and temporal manipulation was found, while both main effects were not significant. This effect seems to be driven by a greater difficulty of the MP condition ($t(16) = 3.20$, $p = 0.006$) (please see supplementary material, Fig. S1C). In this condition, subjects were prioritizing the melody part of a stimulus which had no global temporal leader. This required constant monitoring of the locally fluctuating leader follower roles. No other comparisons were significant.

3.2. Imaging results

3.2.1. Main effects

To explore the neural underpinnings of the two listening styles involved in listening to multi-part music, i.e. segregation and integration, we conducted a 2 × 2 ANOVA with the factors structural relationship and temporal relationship. Detailed results of the 2 × 2 ANOVA are listed in Table 1. First, a main effect of structural relationship yielded significant blood oxygenation level dependent (BOLD) responses in three left hemispheric frontal clusters. Areas included the superior frontal gyrus (SFG), the anterior cingulate cortex (ACC) and the dlPFC (Table 1A; Fig. 2A).

The effect of the temporal relationship showed a significant network of activity within the right dlPFC, right middle temporal gyrus, bilateral IPS and right precuneus as well as in the left inferior parietal lobule (IPL) and the left cerebellum (Table 1B; Fig. 2B).

3.2.2. Interaction (structural relationship x temporal relationship)

Most importantly, the interaction of both leader–follower relationship factors (i.e. structural relationship and temporal relationship between parts) showed a network comprised of right frontal (dlPFC) and right parietal (IPS and IPL) areas (Table 1C; Fig. 3A). Obtained parameter estimates show that this interaction is driven by a bias in percent signal change for attending to melody in the original performance (MP) (Figs. 3B-C) (please see supplementary material for significant imaging contrasts, Fig. S2).

Interestingly, this interaction of our two factors does not represent the bias observed in the behavioral leader–follower relationship or quality rating. These data do show however, a bias for musical structure when attending to melody, which is boosted through the temporal manipulation. This demonstrates that behavioral and imaging data diverge for the factor of temporal manipulation.

Based on our behavioral findings, we included the difficulty rating as a regressor into the 2 × 2 ANOVA design for a second analysis. This does not affect the significant main effects just shown. The interaction of the two factors however is no longer significant (see supplementary material, Table S1A–B, Figs. S1A–B), suggesting an influence of cognitive resources.

4. Discussion

Multi-part music provides an ideal means of looking into forms of attention (Bigand et al., 2000; Keller, 2001; Madsen, 1997) as well as complex perceptual auditory processes (Bregman, 1990; Janata et al., 2002; Satoh et al., 2001). To provide a foundation for future research into how we listen to music and which components capture our attention, perception of musical stimuli, and simple auditory stimuli have recently been investigated (Janata et al., 2002; Madsen, 1997; Pugh et al., 1996; Satoh et al., 2001). As a key aspect of listening to

Table 1
Brain regions that showed significant BOLD activity in the (2 × 2) ANOVA. Cluster thresholding was performed with a Z-threshold of 2.3 and a corrected p-value of <0.01 with a cluster-based correction for multiple comparisons using Gaussian Random Field Theory.

| Anatomical structure	x, y, z coordinates			z-score
(A) Main effect of structural relationship				
L superior frontal gyrus	−6	44	30	4.2
L anterior cingulate cortex	−6	44	12	4.1
L dorsolateral prefrontal cortex	−24	36	28	3.61
(B) Main effect of temporal relationship				
R intraparietal sulcus	42	−48	42	4.88
R precuneus	12	−76	52	4.42
R dorsolateral prefrontal cortex	40	4	54	4.25
	42	32	32	4.17
R middle temporal gyrus	64	−36	−10	3.85
	62	−18	−12	3.61
L intraparietal sulcus	−46	−48	40	4.5
	−38	−58	56	3.52
	−38	−62	34	3.36
L inferior parietal lobule	−30	−66	58	3.13
L cerebellum (crus II)	−12	−80	−30	4.09
L cerebellum (crus I)	−30	−62	−32	3.36
(C) Interaction: structural relationship × temporal relationship				
R dorsolateral prefrontal cortex	48	26	30	4.35
	42	18	36	3.92
R intraparietal sulcus	42	−46	36	3.67
	50	−50	50	3.28
R inferior parietal lobule	48	−60	44	3.5
	36	−72	46	3.24
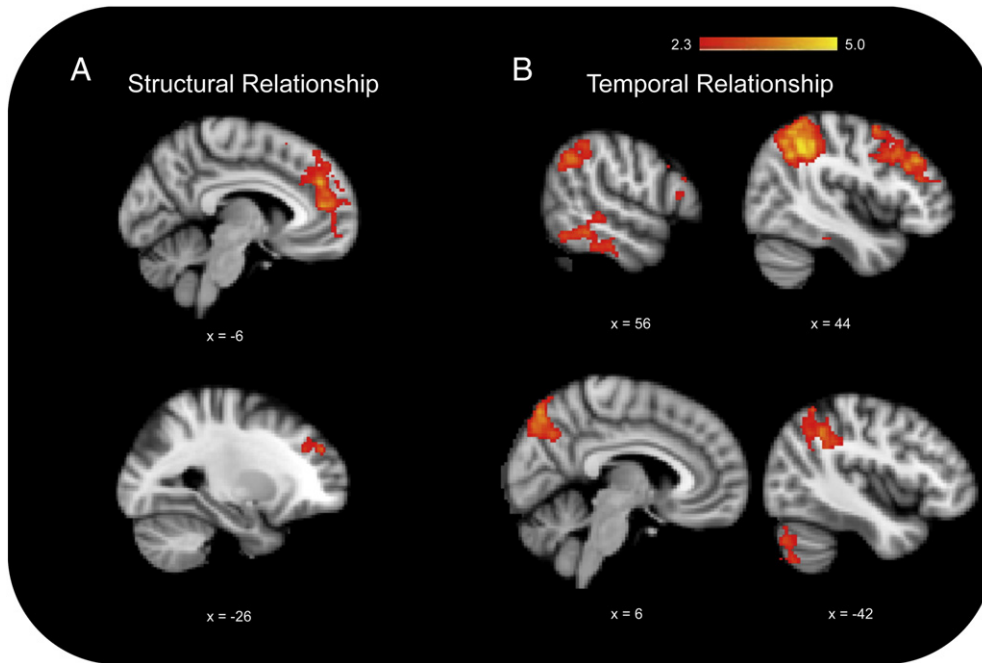
Fig. 2. Main effects of leader–follower relationship factors. Significant ANOVA main effects of (A) structural relationship and (B) temporal relationship between parts. Fixed effects, Z-threshold of 2.3 and a corrected p-value of <0.01.
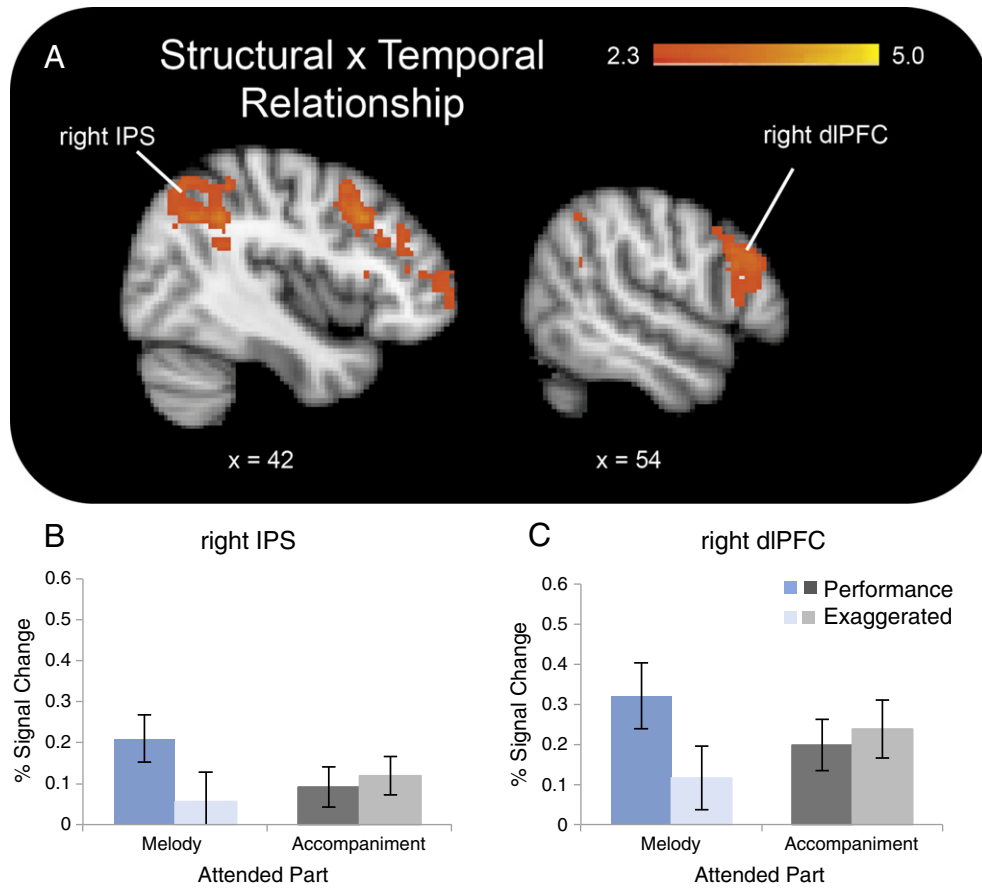

Fig. 3. Interaction of structural and temporal relationship factors. ANOVA interaction. (A) Significant activation clusters for the interaction of structural relationship and temporal relationship revealing a fronto-parietal network of right dorsolateral prefrontal cortex (dlPFC) and the right inferior parietal sulcus (IPS). (B–C) Extracted mean individual parameter estimates for attend melody in performance stimulus (MP) (blue), attend melody in exaggerated stimulus (ME) (light blue), attend accompaniment in performance stimulus (AP) (gray), and attend accompaniment in exaggerated stimulus (AE) (light gray) demonstrating the direction of the corresponding ANOVA interaction activations. Error bars indicate standard error.

multi-part music, we chose to investigate the cognitive basis of the perception and global assessment of a leader–follower relationship between two parts of a piano duet. Based on the recent literature on perception of music (Bigand et al., 2000; Keller and Burnham, 2005), we explored the behavioral as well as neural underpinnings of prioritized integrative attention. More specifically, due to the combination of this specific attention task and a global assessment of the relationship between parts, we investigate how integration and segregation differ in terms of their neural representation. In the following, we will discuss the implications of our results as well as an interaction of saliency and the importance of stream integration in listening to multi-part music.

4.1. Leader–follower relationship influences perception

The assessment of the leader–follower relationship between parts shows a clear interaction of the two relationship factors manipulated in this study. In the attending to melody conditions (i.e. comparing MP with ME), subjects seem to have been basing their assessment primarily on the temporal relationship between parts. The exaggerated (ME) stimuli, containing a global temporal melody lead, were assessed based on the temporal relationship and correctly judged as leading. When looking at the results for the attend to accompaniment conditions (AP & AE) however, we find a structural bias. Here, despite a global temporal lead, the attended stimulus (AE) was not judged as leading.

As integration of the two parts is needed in order to globally assess the relationship between them, we posit that a structural hierarchy biases the integration part of the task. Specifically, our results show that the salient temporal lead of the exaggerated stimulus seems to have been potentiated by the structural salience of the melody (Nothdurft, 2006). This might either be because of a structural dominance the melody has in the particular piece chosen for the study, or because of physical properties which make it more salient. Nevertheless, the interaction of the two factors biased either the cognitive assessment itself or the perception of the relationship between parts and resulted in the melody being judged as leading and the accompaniment as following.

The subjective assessment of the overall quality of a piece seems to also depend on the two manipulated factors. As previously reported for multi-part performances, a temporal melody lead is considered to have a higher quality (Goebl and Palmer, 2009; Keller and Appel, 2010). This hints at a preference for some asynchrony or temporal leader–follower relationship between parts. Both factors might have to interact in a certain fashion in order for the music to be perceived as having a high quality. This could explain the bias we see in the present study for the leader–follower judgment.

4.2. Segregation, integration and the main effects

4.2.1. Segregation
Multi-part music represents a complex auditory stimulus with multiple concurrent streams. When asked to prioritize attention to one part of the duet, subjects need to not only segregate the two concurrent streams but also continue to keep them separate online. This task requires sustained attention akin to monitoring one's own playing in an ensemble (Strait and Kraus, 2011). Musicians accordingly have improved abilities in separating streams (Parbery-Clark et al., 2009; Strait et al., 2010; Zendel and Alain, 2009) and specifically enhanced processing capabilities when it comes to their own instrument (Margulis et al., 2009; Pantev et al., 2001). Presented results for the main effect of structural relationship accordingly show activation indicative of cognitive processes involved in stream segregation. The caudal part of the ACC has extensively been discussed as being involved in task monitoring (Amodio and Frith, 2006; Botvinick et al., 1999; Bush et al., 2000; Carter and van Veen, 2007; Carter et al., 1998; Kerns et al., 2004; Pallesen et al., 2010; Van Veen and Carter, 2002). Commonly in these tasks, the dlPFC is co-activated, leading Carter and Van Veen (2007) to propose that the

ACC becomes activated during task monitoring, when a conflict arises. Two harmonically congruent streams similar in sound, which have to be separated and are competing for resources, could be seen to provide such a conflict. The dlPFC is co-activated to then resolve this conflict and increase attention to the correct stream, increasing cognitive control (Carter and van Veen, 2007). In this way, attention can be prioritized to one stream during continuous conflict. These findings are in line with research showing musicians to activate fronto-parietal working-memory and attention networks to a greater extent than non-musicians in pitch working memory tasks (Gaab et al., 2003; Pallesen et al., 2010). Taken together, the enhanced segregation capabilities of musicians and the BOLD activation shown for our factor of structural relationship fit well with recent literature and point to a top-down monitored segregation process within our task.

4.2.2. Integration
The leader–follower relationship assessment however could not be done solely by segregating the two duet streams and attending to one part. For this and the assessment of quality, it was crucial to prioritize the cued stream while concurrently integrating the second so as to judge the relationship between the streams. The main effect of the temporal relationship between musical parts reflects the integration component of our task.

The role of the middle temporal gyrus in accessing semantic meaning has been discussed extensively in the language literature (e.g. (Hickok and Poeppel, 2000, 2004)). For music perception it has similarly been proposed to reflect a sound-to-meaning interface, processing abstract aspects of music (Grahn and Schuit, 2012; Seung et al., 2005). The integration of different acoustical attributes could thus have guided the decision of a temporal leader or follower within the presented stimulus (Seung et al., 2005).

The IPS, as part of the observed network, has also recently been discussed to be involved with integration of various attributes (Cusack, 2005; Hill and Miller, 2010; Wei et al., 2011). Moreover, other studies in which IPS activity was found concluded similarly that the IPS computes relationships between stimulus elements (Champod and Petrides, 2007; Foster and Zatorre, 2010; Shafritz et al., 2002; Wei et al., 2011; Zatorre et al., 2010) or integrates their neural representations (Alexander et al., 2005; Cusack, 2005; Donner et al., 2002; Hill and Miller, 2010). Together, this more abstract role for the IPS in conjunction the implication of the dlPFC in task monitoring and keeping information in working memory (Champod and Petrides, 2007; Petrides, 2000) keenly describe the processes involved in the assessment of the relationship between musical parts.

Specifically, for the network of the right IPS and the dlPFC identified in the present study, it has been argued that it is involved in encoding and maintaining a representation of time (Alexander et al., 2005; Battelli et al., 2007; Bueti and Walsh, 2009; Janssen and Shadlen, 2005; Koch et al., 2003; Leon and Shadlen, 2003; Rao et al., 2001; VanRullen, 2008; Walsh, 2003). Specifically, the role of the dlPFC within this network may relate to the maintenance of interval representations in working memory as well as their manipulation, as in the comparison of two intervals in order to make a decision about their lengths (Lewis and Miall, 2006; Onoe et al., 2001; Rao et al., 2001; Walsh, 2003). In the present study, the updating of the attended stream's role as leader or follower fits well within this description as well as comparing which of the two streams came first in time. Interestingly, it seems like the right dlPFC in particular is involved in time perception tasks (Koch et al., 2002, 2003). Down-regulating the right dlPFC with repetitive TMS, results in an underestimation of time intervals (Koch et al., 2003). The authors propose that the right dlPFC is part of a timing network in which its function is related to the accumulation of pulses in working memory. The deficit in underestimating time intervals in this study was either due to a slowed down encoding of the accumulated pulses within the network or the decision process when comparing two intervals was disrupted.

Both possibilities point to the involvement of the right dlPFC in particular in time processing.

The other cortical area hypothesized to be involved in time processing is the IPL. Similarly to the former study Alexander and colleagues down-regulated the right and left IPL in turn, to test their involvement in rapid discrimination of temporal intervals. While no effect on task performance was found when the left IPL was stimulated, repetitive TMS over the right IPL interfered with participants' ability to judge temporal intervals in both hemifields (Alexander et al., 2005). The authors concluded that taken together with IPL activation in other tasks, this structure may incorporate representations that are common to space, time and quantity (Alexander et al., 2005; Bueti and Walsh, 2009; Walsh, 2003). A recent study on temporal order judgments additionally suggests that the right IPL is central for such judgments or more generally tasks that depend on the control of attention over time (Battelli et al., 2007). Besides these human functional imaging and brain stimulation studies, single cell recordings within the lateral intraparietal area (LIP) of the intraparietal sulcus in the macaque monkey show that neurons within this area represent elapsed time (Janssen and Shadlen, 2005; Leon and Shadlen, 2003). Involvement of the IPS thus corresponds well with the correct assessment of the temporal manipulation in the leader–follower relationship observed in the behavioral data.

An extended neural network comprising not only the dlPFC and IPS but also cerebellar and basal ganglia activity for timing dependent tasks has also already been reported (Jantzen et al., 2007; O'Reilly et al., 2008; Rao et al., 2001; Schubotz et al., 2000; Smith et al., 2003; Thaut et al., 2009; Tracy et al., 2000; Wiener et al., 2010). Moreover, a recent functional connectivity study on cerebellar-frontal circuits revealed strong correlations in connectivity for crus I and crus II, activated in the present study with the medial prefrontal cortex (PFC) and dlPFC respectively (Krienen and Buckner, 2009). This is in line with studies arguing that the dlPFC is involved not only in cognitive control and working memory tasks (Carter and van Veen, 2007; Pallesen et al., 2010) but also in timing (Rao et al., 2001; Smith et al., 2003). Taken together the BOLD activation within the right IPS and the dlPFC for the temporal relationship factor fits well within the recent literature and suggests the involvement of these areas in both segregation and integration as well as the assessment of the temporal relationship between parts in a musical duet.

4.3. Interaction between factors and saliency

Both behaviorally and neutrally, we observe a so-called structural bias. Our results indicate an influence of both the temporal and structural relationship between parts. More interestingly, our data highlight an interaction of our two factors such that the structural bias is heightened by the temporal relationship. We posit that it is an interaction of the saliency of both factors which drives this effect resulting in the observed bias of the subjective assessment and the neural activity in a fronto-parietal attention network. This interaction was specifically induced due to the employed task in which subjects had to integrate both parts to get at the relationship between them (Bigand et al., 2000; Keller, 2001, 2008). It is interesting to note, however, that the effect of salience is different at the behavioral and neural level.

4.3.1. Salience

The salience of the structural and the temporal relationship is reflected directly in the subjective assessment of the leader–follower relationship. Only the melody (and not the accompaniment) was rated as leading when it was temporally leading, suggestive of the fact that the relationship factors must interact on a perceptual level. The imaging results, on the other hand, show the inverse bias of the structural relationship. The condition in which attention was prioritized to the melody stream that was not globally temporally ahead (MP) shows the highest percent signal change. We hypothesize that

this effect is due to an interaction of the saliency of the two relationship factors. Not only is the melody part structurally more dominant (Bregman, 1990) but has often, and most certainly in the music we chose, more salient physical properties. Except for choir singers and ensemble musicians who are used to producing and following the accompaniment, listeners tend to prioritize and attend to the melody of a piece (Madsen, 1997). In general, individuals are much more used to singing and remembering the melody of a multi-part musical piece than the accompaniment, which could increase the melodic salience (Jagadeesh et al., 2001). Nevertheless, an increase in salience could also be due to general *gestalt* principles and grouping mechanisms (Bregman, 1990; Drake et al., 2000). As participants were forced to group the music they listened to in a horizontal as opposed to a vertical fashion, the principles for grouping melody might have been much stronger and therefore more salient (Tse, 2005). Moreover, a study on finger tapping to expressively played music found that musicians organize auditory events over a longer time span and focus on higher hierarchical levels within metric sequences (Drake et al., 2000). The authors concluded that, overall, musicians have a better sense of the hierarchical structure of a piece than non-musicians do. In the present study, specific qualities such as the higher pitch range of the melody could have increased its overall salience (McAdams and Drake, 2002). The salience bias observed in this study in favor of the melody may thus be confounded with the higher pitch range and familiarity. Future studies looking into musical pieces with an alternating melody between different parts could shed light on the impact of the frequency range. We suggest that salience of the structural relationship is the dominant factor, which can be modulated by the additional salience of the specific frequency range.

4.3.2. Cognitive load

The observed effect of salience was increased by the cued prioritization of the attended part during listening. Single cell recordings as well as fMRI studies have shown attention to bias neural responses by increasing the attended stimulus' salience (Reddy et al., 2009; Reynolds and Desimone, 2003). With the highly salient melody, it would have been very difficult to concurrently integrate the much less salient accompaniment (Nothdurft, 2006; Reddy et al., 2009; Reynolds and Desimone, 2003) as reflected by the ratings of perceived difficulty for the MP condition. The other attend to melody condition (ME) does not show the same BOLD effect, which one would expect if only structural salience was the driving factor.

We therefore suggest the additional influence of difficulty or cognitive load (Adler et al., 2001; Pugh et al., 1996). This is supported by the fact that the interaction fails to become significant, when controlling for perceived difficulty (see supplementary material). Interference during attention tasks has been shown to increase with cognitive load (Lavie and De Fockert, 2005; Lavie et al., 2004). In our task, interference was greatest when attention was captured by the prioritized and more salient melody in the performance without a global leader (MP). The nature of our instructed task was such that subjects had to prioritize one stream and integrate the second, while additionally having to continuously assess the relationship between parts. The latter of these tasks required more cognitive resources for the performance stimulus in which the temporal relationship only varied locally and as such no obvious global (temporal) leader could be identified. This stimulus necessitated constant monitoring of the temporal relationship, resulting in increased cognitive load. This potentiation of cognitive load by salience and monitoring could explain the higher BOLD response as a result of the interaction of the two leader–follower relationship factors. It further leads to the conclusion that the integration part of the task rather than the segregation part was the one in which both relationship factors interacted.

4.3.3. Fronto-parietal network

Support for this claim comes from studies in the visual domain which discuss this fronto-parietal network in relation to selective attention and feature integration (Donner et al., 2002; Shafritz et al.,

2002; Wei et al., 2011). Observed activation of these areas however has not been restricted to visual attention tasks. Spatial transformations such as mental rotation as well as pitch or melody transformations have been shown to elicit activation in the PFC and the IPS (Foster and Zatorre, 2010; Harris and Miniussi, 2003; Zatorre et al., 2010). Further evidence for modality independent processing within the IPS stems from research on macaque brains and corresponding human analog functional imaging data (for a detailed review see (Grefkes and Fink, 2005)). A diffusion tensor magnetic resonance imaging study on humans additionally found that these fronto-parietal areas are linked through white matter fiber pathways (Makris et al., 2005).

Beyond the described role of the IPS and PFC in the integration of attributes or a representation of time, others have implicated these structures in top-down attentional modulation of stimuli (Bressler et al., 2008; Corbetta and Shulman, 2002; Cusack, 2005; Davranche et al., 2011; Marois et al., 2004; Wei et al., 2011). Recent studies from the visual domain report top-down modulatory effects to be reflected by an increase in BOLD response (Eger et al., 2007; Gazzaley et al., 2005). Specifically, it has been suggested that these modulations may function by increasing stimulus salience in a top-down fashion (Nothdurft, 2006; Reddy et al., 2009; Reynolds and Desimone, 2003).

5. Conclusion

In the present study, we present a novel paradigm for exploring so-called prioritized integrative attention, a kind of attention used in playing and listening to music. In particular this paradigm probes segregation and integration mechanisms involved in perceiving complex auditory scenes, such as multi-part music. It is thus an innovative way of exploring the behavioral and neural underpinnings of the assessment of relationships between concurrent auditory streams. We identify a right sided fronto-parietal network for the interaction of the two manipulated leader–follower relationship factors that most likely allows for the integration and the segregation of streams, calculating the relationship between them and biasing perception through top-down salience.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neuroimage.2013.03.051.

Conlict of interest

The authors declare no conflict of interest.

References

Adler, C.M., Sax, K.W., Holland, S.K., Schmithorst, V., Rosenberg, L., Strakowski, S.M., 2001. Changes in neuronal activation with increasing attention demand in healthy volunteers: an fMRI study. Synapse 42, 266–272.
Alexander, I., Cowey, A., Walsh, V., 2005. The right parietal cortex and time perception: back to Critchley and the Zeitraffer phenomenon. Cogn. Neuropsychol. 22, 306–315.
Amodio, D.M., Frith, C.D., 2006. Meeting of minds: the medial frontal cortex and social cognition. Nat. Rev. Neurosci. 7, 268–277.
Battelli, L., Pascual-Leone, A., Cavanagh, P., 2007. The 'when' pathway of the right parietal lobe. Trends Cogn. Sci. 11, 204–210.
Bigand, E., McAdams, S., Forêt, S., 2000. Divided attention in music. Int. J. Psychol. 35, 270–278.
Botvinick, M., Nystrom, L.E., Fissell, K., Carter, C.S., Cohen, J.D., 1999. Conflict monitoring versus selection-for-action in anterior cingulate cortex. Nature 402, 179–181.
Bregman, A.S., 1990. Auditory Scene Analysis: The Perceptual Organization of Sound. MIT Press, Cambridge, MA.
Bressler, S.L., Tang, W., Sylvester, C.M., Shulman, G.L., Corbetta, M., 2008. Top-down control of human visual cortex by frontal and parietal cortex in anticipatory visual spatial attention. J. Neurosci. 28, 10056–10061.
Bueti, D., Walsh, V., 2009. The parietal cortex and the representation of time, space, number and other magnitudes. Philos. Trans. R. Soc. Lond. B Biol. Sci. 364, 1831–1840.
Bush, G., Luu, P., Posner, M.I., 2000. Cognitive and emotional influences in anterior cingulate cortex. Trends Cogn. Sci. 4, 215–222.
Carter, C.S., van Veen, V., 2007. Anterior cingulate cortex and conflict detection: an update of theory and data. Cogn. Affect. Behav. Neurosci. 7, 367–379.
Carter, C.S., Braver, T.S., Barch, D.M., Botvinick, M.M., Noll, D., Cohen, J.D., 1998. Anterior cingulate cortex, error detection, and the online monitoring of performance. Science 280, 747–749.

Champod, A.S., Petrides, M., 2007. Dissociable roles of the posterior parietal and the prefrontal cortex in manipulation and monitoring processes. Proc. Natl. Acad. Sci. U. S. A. 104, 14837–14842.
Corbetta, M., Shulman, G.L., 2002. Control of goal-directed and stimulus-driven attention in the brain. Nat. Rev. Neurosci. 3, 201–215.
Cusack, R., 2005. The intraparietal sulcus and perceptual organization. J. Cogn. Neurosci. 17, 641–651.
Davranche, K., Nazarian, B., Vidal, F., Coull, J., 2011. Orienting attention in time activates left intraparietal sulcus for both perceptual and motor task goals. J. Cogn. Neurosci. 3318–3330.
Donner, T.H., Kettermann, A., Diesch, E., Ostendorf, F., Villringer, A., Brandt, S.A., 2002. Visual feature and conjunction searches of equal difficulty engage only partially overlapping frontoparietal networks. NeuroImage 15, 16–25.
Drake, C., Penel, A., Bigand, E., 2000. Tapping in time with mechanically and expressively performed music. Music. Percept. 18, 1–23.
Eger, E., Henson, R.N., Driver, J., Dolan, R.J., 2007. Mechanisms of top-down facilitation in perception of visual objects studied by FMRI. Cereb. Cortex 17, 2123–2133.
Erickson, R., 1975. Sound Structure in Music. University of California Press, Berkley and Los Angeles, CA.
Foster, N.E., Zatorre, R.J., 2010. A role for the intraparietal sulcus in transforming musical pitch information. Cereb. Cortex 20, 1350–1359.
Friston, K.J., Jezzard, P., Turner, R., 1994. Analysis of functional MRI time-series. Hum. Brain Mapp. 153–171.
Gaab, N., Gaser, C., Zaehle, T., Jancke, L., Schlaug, G., 2003. Functional anatomy of pitch memory — an fMRI study with sparse temporal sampling. NeuroImage 19, 1417–1426.
Gazzaley, A., Cooney, J.W., McEvoy, K., Knight, R.T., D'Esposito, M., 2005. Top-down enhancement and suppression of the magnitude and speed of neural activity. J. Cogn. Neurosci. 17, 507–517.
Goebl, W., Palmer, C., 2009. Synchronization of timing and motion among performing musicians. Music. Percept. 26, 427–438.
Grahn, J.A., Schuit, D., 2012. Individual differences in rhythmic ability: behavioral and neuroimaging investigations. Psychomusicology 22, 105–121.
Grefkes, C., Fink, G.R., 2005. The functional organization of the intraparietal sulcus in humans and monkeys. J. Anat. 207, 3–17.
Handel, S., 1989. Listening: An Introduction to the Perception of Auditory Events. The MIT Press, Cambridge, MA, US.
Harris, I.M., Miniussi, C., 2003. Parietal lobe contribution to mental rotation demonstrated with rTMS. J. Cogn. Neurosci. 15, 315–323.
Hickok, G., Poeppel, D., 2000. Towards a functional neuroanatomy of speech perception. Trends Cogn. Sci. 4, 131–138.
Hickok, G., Poeppel, D., 2004. Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. Cognition 92, 67–99.
Hill, K.T., Miller, L.M., 2010. Auditory attentional control and selection during cocktail party listening. Cereb. Cortex 20, 583–590.
Jagadeesh, B., Chelazzi, L., Mishkin, M., Desimone, R., 2001. Learning increases stimulus salience in anterior inferior temporal cortex of the macaque. J. Neurophysiol. 86, 290–303.
Janata, P., Tillmann, B., Bharucha, J.J., 2002. Listening to polyphonic music recruits domain-general attention and working memory circuits. Cogn. Affect. Behav. Neurosci. 2, 121–140.
Janssen, P., Shadlen, M.N., 2005. A representation of the hazard rate of elapsed time in macaque area LIP. Nat. Neurosci. 8, 234–241.
Jantzen, K.J., Oullier, O., Marshall, M., Steinberg, F.L., Kelso, J.A., 2007. A parametric fMRI investigation of context effects in sensorimotor timing and coordination. Neuropsychologia 45, 673–684.
Jenkinson, M., Smith, S., 2001. A global optimisation method for robust affine registration of brain images. Med. Image Anal. 5, 143–156.
Keller, P.E., 2001. Attentional resource allocation in musical ensemble performance. Psychol. Music. 29, 20–38.
Keller, P.E., 2008. Joint action in music performance. In: Morganti, F., Carassa, A., Riva, G. (Eds.), Enacting Intersubjectivity: A Cognitive and Social Perspective to the Study of Interactions. IOS Press, Amsterdam, pp. 205–221.
Keller, P.E., Appel, M., 2010. Individual differences, auditory imagery, and the coordination of body movements and sounds in musical ensembles. Music. Percept. 28, 27–46.
Keller, P.E., Burnham, D.K., 2005. Musical meter in attention to multipart rhythm. Music. Percept. 22, 629–661.
Kerns, J.G., Cohen, J.D., MacDonald III, A.W., Cho, R.Y., Stenger, V.A., Carter, C.S., 2004. Anterior cingulate conflict monitoring and adjustments in control. Science 303, 1023–1026.
Koch, G., Oliveri, M., Carlesimo, G.A., Caltagirone, C., 2002. Selective deficit of time perception in a patient with right prefrontal cortex lesion. Neurology 59, 1658–1659.
Koch, G., Oliveri, M., Torriero, S., Caltagirone, C., 2003. Underestimation of time perception after repetitive transcranial magnetic stimulation. Neurology 60, 1844–1846.
Krienen, F.M., Buckner, R.L., 2009. Segregated fronto-cerebellar circuits revealed by intrinsic functional connectivity. Cereb. Cortex 19, 2485–2497.
Lavie, N., De Fockert, J., 2005. The role of working memory in attentional capture. Psychon. Bull. Rev. 12, 669–674.
Lavie, N., Hirst, A., de Fockert, J.W., Viding, E., 2004. Load theory of selective attention and cognitive control. J. Exp. Psychol. Gen. 133, 339–354.
Leon, M.I., Shadlen, M.N., 2003. Representation of time by neurons in the posterior parietal cortex of the macaque. Neuron 38, 317–327.
Lewis, P.A., Miall, R.C., 2006. Remembering the time: a continuous clock. Trends Cogn. Sci. 10, 401–406.
Madsen, C.K., 1997. Focus of attention and aesthetic response. J. Res. Music. Educ. 45, 80–89.
Makris, N., Kennedy, D.N., McInerney, S., Sorensen, A.G., Wang, R., Caviness Jr., V.S., Pandya, D.N., 2005. Segmentation of subcomponents within the superior

longitudinal fascicle in humans: a quantitative, in vivo, DT-MRI study. Cereb. Cortex 15, 854–869.

Margulis, E.H., Mlsna, L.M., Uppunda, A.K., Parrish, T.B., Wong, P.C., 2009. Selective neurophysiologic responses to music in instrumentalists with different listening biographies. Hum. Brain Mapp. 30, 267–275.

Marois, R., Chun, M.M., Gore, J.C., 2004. A common parieto-frontal network is recruited under both low visibility and high perceptual interference conditions. J. Neurophysiol. 92, 2985–2992.

McAdams, S., Drake, C., 2002. Auditory Perception and Cognition, 3rd ed. John Wiley & Sons, Inc., New York.

Nelken, I., 2011. Music and the auditory brain: where is the connection? Front. Hum. Neurosci. 5, 106.

Nothdurft, H.-C., 2006. Salience and target selection in visual search. Vis. Cogn. 14, 514–542.

Onoe, H., Komori, M., Onoe, K., Takechi, H., Tsukada, H., Watanabe, Y., 2001. Cortical networks recruited for time perception: a monkey positron emission tomography (PET) study. NeuroImage 13, 37–45.

O'Reilly, J.X., Mesulam, M.M., Nobre, A.C., 2008. The cerebellum predicts the timing of perceptual events. J. Neurosci. 28, 2252–2260.

Pallesen, K.J., Brattico, E., Bailey, C.J., Korvenoja, A., Koivisto, J., Gjedde, A., Carlson, S., 2010. Cognitive control in auditory working memory is enhanced in musicians. PLoS One 5, e11120.

Palmer, C., 1997. Music performance. Annu. Rev. Psychol. 48, 115–138.

Pantev, C., Roberts, L.E., Schulz, M., Engelien, A., Ross, B., 2001. Timbre-specific enhancement of auditory cortical representations in musicians. NeuroReport 12, 169–174.

Parbery-Clark, A., Skoe, E., Kraus, N., 2009. Musical experience limits the degradative effects of background noise on the neural processing of sound. J. Neurosci. 29, 14100–14107.

Petrides, M., 2000. The role of the mid-dorsolateral prefrontal cortex in working memory. Exp. Brain Res. 133, 44–54.

Pugh, K.R., offywitz, B.A., Shaywitz, S.E., Fulbright, R.K., Byrd, D., Skudlarski, P., Shankweiler, D.P., Katz, L., Constable, R.T., Fletcher, J., Lacadie, C., Marchione, K., Gore, J.C., 1996. Auditory selective attention: an fMRI investigation. NeuroImage 4, 159–173.

Rao, S.M., Mayer, A.R., Harrington, D.L., 2001. The evolution of brain activation during temporal processing. Nat. Neurosci. 4, 317–323.

Rasch, R.A., 1979. Synchronization in performed ensemble music. Acustica 43, 121–131.

Rasch, R.A., 2000. Timing and synchronization in ensemble performance. In: Sloboda, J.A. (Ed.), Generative Processes in Music. Oxford University Press, Oxford.

Reddy, L., Kanwisher, N.G., VanRullen, R., 2009. Attention and biased competition in multi-voxel object representations. Proc. Natl. Acad. Sci. U. S. A. 106, 21447–21452.

Repp, B.H., 1996. Patterns of note onset asynchronies in expressive piano performance. J. Acoust. Soc. Am. 100, 3917–3932.

Reynolds, J.H., Desimone, R., 2003. Interacting roles of attention and visual salience in V4. Neuron 37, 853–863.

Satoh, M., Takeda, K., Nagata, K., Hatazawa, J., Kuzuhara, S., 2001. Activated brain regions in musicians during an ensemble: a PET study. Brain Res. Cogn. Brain Res. 12, 101–108.

Schubotz, R.I., Friederici, A.D., von Cramon, D.Y., 2000. Time perception and motor timing: a common cortical and subcortical basis revealed by fMRI. NeuroImage 11, 1–12.

Seung, Y., Kyong, J.S., Woo, S.H., Lee, B.T., Lee, K.M., 2005. Brain activation during music listening in individuals with or without prior music training. Neurosci. Res. 52, 323–329.

Shafritz, K.M., Gore, J.C., Marois, R., 2002. The role of the parietal cortex in visual feature binding. Proc. Natl. Acad. Sci. U. S. A. 99, 10917–10922.

Smith, S.M., 2002. Fast robust automated brain extraction. Hum. Brain Mapp. 17, 143–155.

Smith, A., Taylor, E., Lidzba, K., Rubia, K., 2003. A right hemispheric frontocerebellar network for time discrimination of several hundreds of milliseconds. NeuroImage 20, 344–350.

Strait, D.L., Kraus, N., 2011. Can you hear me now? Musical training shapes functional brain networks for selective auditory attention and hearing speech in noise. Front. Psychol. 2, 113.

Strait, D.L., Kraus, N., Parbery-Clark, A., Ashley, R., 2010. Musical experience shapes top-down auditory mechanisms: evidence from masking and auditory attention performance. Hear. Res. 261, 22–29.

Tagg, P., 2003a. Accompaniment. In: Horn, David, Laing, Dave, Oliver, Paul, Wicke, P. (Eds.), Continuum Encyclopedia of Popular Music of the World. Continuum, New York, pp. 628–630.

Tagg, P., 2003b. Melody. In: Horn, David, Laing, Dave, Oliver, Paul, Wicke, P. (Eds.), Continuum Encyclopedia of Popular Music of the World. Continuum, New York, pp. 567–587.

Thaut, M.H., Stephan, K.M., Wunderlich, G., Schicks, W., Tellmann, L., Herzog, H., McIntosh, G.C., Seitz, R.J., Homberg, V., 2009. Distinct cortico-cerebellar activations in rhythmic auditory motor synchronization. Cortex 45, 44–53.

Tracy, J.I., Faro, S.H., Mohamed, F.B., Pinsk, M., Pinus, A., 2000. Functional localization of a "time keeper" function separate from attentional resources and task strategy. NeuroImage 11, 228–242.

Tse, P.U., 2005. Voluntary attention modulates the brightness of overlapping transparent surfaces. Vision Res. 45, 1095–1098.

Van Veen, V., Carter, C.S., 2002. The timing of action-monitoring processes in the anterior cingulate cortex. J. Cogn. Neurosci. 14, 593–602.

VanRullen, R., 2008. The continuous wagon wheel illusion and the 'when' pathway of the right parietal lobe: a repetitive transcranial magnetic stimulation study. PLoS ONE 3 (8), e2911.

Walsh, V., 2003. A theory of magnitude: common cortical metrics of time, space and quantity. Trends Cogn. Sci. 7, 483–488.

Wei, P., Muller, H.J., Pollmann, S., Zhou, X., 2011. Neural correlates of binding features within- or cross-dimensions in visual conjunction search: an fMRI study. NeuroImage 57, 235–241.

Wiener, M., Turkeltaub, P., Coslett, H.B., 2010. The image of time: a voxel-wise meta-analysis. NeuroImage 49, 1728–1740.

Woolrich, M.W., Ripley, B.D., Brady, M., Smith, S.M., 2001. Temporal autocorrelation in univariate linear modeling of FMRI data. NeuroImage 14, 1370–1386.

Worsley, K.J., Evans, A.C., Marrett, S., Neelin, P., 1992. A three-dimensional statistical analysis for CBF activation studies in human brain. J. Cereb. Blood Flow Metab. 12, 900–918.

Wright, J.K., Bregman, A.S., 1987. Auditory stream segregation and the control of dissonance in polyphonic music. Contemp. Music. Rev. 2, 63–92.

Zatorre, R.J., Halpern, A.R., Bouffard, M., 2010. Mental reversal of imagined melodies: a role for the posterior parietal cortex. J. Cogn. Neurosci. 22, 775–789.

Zendel, B.R., Alain, C., 2009. Concurrent sound segregation is enhanced in musicians. J. Cogn. Neurosci. 21, 1488–1498.