



HAL
open science

A functional magnetic resonance imaging examination of audiovisual observation of a point-light string quartet using intersubject correlation and physical feature analysis

Amanda Lillywhite, Dewy Nijhof, Donald Glowinski, Bruno L Giordano, Antonio Camurri, Ian Cross, Frank Pollick

► To cite this version:

Amanda Lillywhite, Dewy Nijhof, Donald Glowinski, Bruno L Giordano, Antonio Camurri, et al.. A functional magnetic resonance imaging examination of audiovisual observation of a point-light string quartet using intersubject correlation and physical feature analysis. *Frontiers in Neuroscience*, 2022, 16, pp.921489. 10.3389/fnins.2022.921489 . hal-03845716

HAL Id: hal-03845716

<https://hal.science/hal-03845716v1>

Submitted on 9 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



OPEN ACCESS

EDITED BY

Javier Gonzalez-Castillo,
National Institute of Mental Health
(NIH), United States

REVIEWED BY

Juha Pajula,
VTT Technical Research Centre
of Finland Ltd., Finland
Anna Czepiel,
Max Planck Institute for Empirical
Aesthetics, Germany

*CORRESPONDENCE

Frank E. Pollick
frank.pollick@glasgow.ac.uk

†These authors share first authorship

SPECIALTY SECTION

This article was submitted to
Brain Imaging Methods,
a section of the journal
Frontiers in Neuroscience

RECEIVED 15 April 2022

ACCEPTED 05 August 2022

PUBLISHED 06 September 2022

CITATION

Lillywhite A, Nijhof D, Glowinski D,
Giordano BL, Camurri A, Cross I and
Pollick FE (2022) A functional magnetic
resonance imaging examination
of audiovisual observation of a
point-light string quartet using
intersubject correlation and physical
feature analysis.
Front. Neurosci. 16:921489.
doi: 10.3389/fnins.2022.921489

COPYRIGHT

© 2022 Lillywhite, Nijhof, Glowinski,
Giordano, Camurri, Cross and Pollick.
This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s)
are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

A functional magnetic resonance imaging examination of audiovisual observation of a point-light string quartet using intersubject correlation and physical feature analysis

Amanda Lillywhite^{1,2†}, Dewy Nijhof^{1,3†}, Donald Glowinski^{4,5†},
Bruno L. Giordano⁶, Antonio Camurri⁷, Ian Cross⁸ and
Frank E. Pollick^{1*}

¹School of Psychology & Neuroscience, University of Glasgow, Glasgow, United Kingdom, ²Department of Psychology, University of Bath, Bath, United Kingdom, ³Institute of Health & Wellbeing, University of Glasgow, Glasgow, United Kingdom, ⁴La Source School of Nursing, Institut et Haute Ecole de la Santé La Source (HES-SO), Lausanne, Switzerland, ⁵Swiss Center for Affective Sciences, University of Geneva, Geneva, Switzerland, ⁶Institut de Neurosciences de la Timone, UMR 7289, CNRS, Aix-Marseille University, Marseille, France, ⁷Casa Paganini-InfoMus, DIBRIS, University of Genoa, Genoa, Italy, ⁸Centre for Music and Science, Faculty of Music, School of Arts and Humanities, University of Cambridge, Cambridge, United Kingdom

We use functional Magnetic Resonance Imaging (fMRI) to explore synchronized neural responses between observers of audiovisual presentation of a string quartet performance during free viewing. Audio presentation was accompanied by visual presentation of the string quartet as stick figures observed from a static viewpoint. Brain data from 18 musical novices were obtained during audiovisual presentation of a 116 s performance of the allegro of String Quartet, No. 14 in D minor by Schubert played by the 'Quartetto di Cremona.' These data were analyzed using intersubject correlation (ISC). Results showed extensive ISC in auditory and visual areas as well as parietal cortex, frontal cortex and subcortical areas including the medial geniculate and basal ganglia (putamen). These results from a single fixed viewpoint of multiple musicians are greater than previous reports of ISC from unstructured group activity but are broadly consistent with related research that used ISC to explore listening to music or watching solo dance. A feature analysis examining the relationship between brain activity and physical features of the auditory and visual signals yielded findings of a large proportion of activity related to auditory and visual processing, particularly in the superior temporal gyrus (STG) as well as midbrain areas. Motor areas were also involved, potentially as a result of watching motion from the stick figure display of musicians in the string quartet. These results reveal

involvement of areas such as the putamen in processing complex musical performance and highlight the potential of using brief naturalistic stimuli to localize distinct brain areas and elucidate potential mechanisms underlying multisensory integration.

KEYWORDS

music, biological motion, fMRI, intersubject correlation (ISC), string quartet

Introduction

Observing coordinated action is a complicated task that typically involves processing the relationship between what is seen and what is heard as well as the social relationship between the people being observed. Moreover, it requires this perceptual and social processing to keep up with the ongoing activity. While much is known about the constituent neural mechanisms that might subservise these processes (Grosbras et al., 2012; de Gelder and Solanas, 2021; Schurz et al., 2021), less is known about how the brain responds to naturalistic stimuli that depict coordinated activity of a group of individuals. In this research we use a string quartet as an example of coordinated human activity and examine, with a data-driven exploratory functional Magnetic Resonance Imaging (fMRI) study, the response of musical novices. Besides its relevance to the study of music neuroscience, a musical activity was chosen since ensemble music is widespread and is one of the most closely synchronized human activities and thus provides a relevant and well-defined activity. Musical novice observers were chosen since some studies suggest that experts may employ mechanisms for integrating the sights and sounds of musical activities that are not typical of the general population (Pettrini et al., 2011).

Vision has been shown to play an important role in the processing of musical performance, with a meta-analysis of 15 studies of audiovisual music perception revealing a clear influence of vision (Platz and Kopiez, 2012). Ratings such as expressiveness (Wanderley et al., 2005; Dahl and Friberg, 2007; Vines et al., 2011), liking or overall quality of performance differed between audio-only and audiovisual presentation. This finding is consistent with claims that vision will interact with the processing of sound at multiple levels of music processing (Schutz, 2008). One example of an interaction between vision and sound production is provided by evidence that musicians playing solo or in ensemble move differently depending on visibility of the other players (Keller and Appel, 2010; Palmer, 2012). A possible reason for these differences in movement arise from the way that different sensory information interacts with basic timing mechanisms that facilitate coordinated play (Wing et al., 2014). Ensemble performance is thought to arise from entrainment of the musicians (Phillips-Silver and Keller, 2012) to a rhythmic signal (Schachner et al., 2009). Another

visual aspect of ensemble playing is the social communication among players afforded by gesture (Glowinski, 2013; Glowinski et al., 2014) that can signal aspects of control, approval and other information. Interaction between biological motion and audiovisual features has been demonstrated in reaction times for detecting visual biological motion, where visual biological motion embedded in noise was detected faster if there was congruent auditory motion (Brooks et al., 2007). Studying this phenomenon in naturalistic stimuli, namely tap dancing, Arrighi et al. (2009) showed that auditory and visual signals integrate in the perception of biological motion only if visual and auditory sequences are synchronized. Moreover, modulation of biological motion perception by kinematic sonification has been shown in gross motor cyclic movements, which were perceived as faster or slower depending on hearing higher or lower global pitch, respectively (Effenberg and Schmitz, 2018). Regardless of the origin of movement and sound differences in the production of ensemble playing, observers are presented with a multisensory signal that is shaped by constraints of ensemble play. In the current research we investigate how a novice group of observers process an ensemble performance by examining brain activity as revealed by fMRI while participants watch and listen to a carefully controlled video of a string quartet ensemble performance (Glowinski, 2013). Brain structures have been shown to entrain to the beat and meter of a musical signal (Iversen et al., 2009; Nozaradan et al., 2011; Fujioka et al., 2012), and the current study informs our understanding of how the audiovisual signal produced by ensemble performance is processed.

Many different auditory features exist in musical stimuli, some of which can be actively distinguished, such as pitch or rhythm and some of which often cannot, such as harmonics. Whilst the term auditory features embodies a multitude of concepts at various levels whose combined effect is larger than its individual components, for the purpose of this study we refer to auditory features as preconceived characteristics of the auditory signal. Although neural correlates of auditory features have often been studied using neuroimaging, neural processing of a continuous stream of auditory features is relatively under examined. Loudness is an often-investigated feature, measured in dB of sound pressure level (SPL), although cortical activations seem to relate to perceived levels of loudness as opposed to

the physical sound intensity (Langers et al., 2007; Röhl and Uppenkamp, 2012). Perceptual levels of loudness are therefore measured in sones. For loudness, activation is usually found in the primary auditory cortex (Röhl et al., 2011). Other commonly investigated auditory features include pitch and spatial processing. A meta-analysis by Alho et al. (2014) found activation in the superior temporal gyrus (STG) across 115 fMRI studies that attempted to localize the functions of pitch and spatial location within the auditory cortex. Giordano et al. (2013) found that large portions of the bilateral temporal cortex were related to low level feature processing (pitch, spectral centroid, harmonicity, and loudness). This included auditory areas such as Heschl's gyrus, the planum temporale (PT) and the anterior STG. In terms of processing, Chapin et al. (2010) found that several areas of cortical activity were correlated with tempo in expressive performances.

A large body of neuroimaging work has correlated different low-level auditory features to specific cortical and subcortical activation using a classic General Linear Model (GLM) design with simple, short stimuli. These studies are able to provide valuable information about the basic mechanisms that underlie these processes, as much of the interference from other factors are excluded to provide more precise and reliable results. However, at the same time the act of excluding this interference limits the extent of knowledge they can provide on natural perceptual processes, whether they are auditory, visual or multimodal. In typical everyday life, scenes are often more complex and, with the inclusion of many objects in a scene as well as the presence of more than one sensory modality, require multisensory integration. This drives research into the use of more complex and naturalistic stimuli.

A considerable amount of research investigating perception with more complex extended stimuli has emerged. However, the use of these complex extended stimuli can be problematic with typical GLM analyses, as GLM analyses typically place constraints on the kind of stimuli that can be analyzed. A significant challenge to studying brain mechanisms of audiovisual perception of a string quartet is how to design the acquisition and analysis of brain data that preserves the intrinsic properties of the string quartet performance. Our solution to this was to use the technique of Intersubject Correlation (ISC) which enables examination of brain activity to complex stimuli extended over minutes (Hasson et al., 2004). ISC is a model-free approach that measures how, among a set of observers, activity at different regions of the brain is correlated among the observers. Evidence has emerged showing the role of correlated activity amongst observers, on a neural, behavioral and physiological level, in shared experience of a musical performance (e.g., Ardizzi et al., 2020; Hou et al., 2020). In this view, ISC could be interpreted as an index of coordinated audience response. This ISC can be extensive as demonstrated by the original study of Hasson et al. (2004) which showed that when a group of observers watched an excerpt of Sergio Leone's film *The Good, the Bad and the Ugly* (1966), ISC was

found in about 45% of cortex. One advantage of ISC is that it is independent of the level of activity within the brain, which circumvents the issue that the hemodynamic response function varies not only between brain areas, but also between subjects (Jääskeläinen et al., 2008). It has also been shown across a variety of studies to be a reliable method to investigate neural activity during complex stimuli, when compared to classic modeling approaches using the GLM (Pajula et al., 2012). ISC has been used with fMRI data to study a variety of topics including memory (Hasson et al., 2008a), perception of film (Hasson et al., 2008b), temporal receptive windows (Hasson et al., 2008c) and communication (Hasson et al., 2012; Rowland et al., 2018). It has also been applied to the study of group differences due to development (Cantlon and Li, 2013), autism (Hasson et al., 2009) and the expertise of CCTV operators in judging harmful intent (Petrini et al., 2014). Finally, the ISC approach has also been adapted for use in electroencephalogram (EEG) (Dmochowski et al., 2012) and functional near infrared spectroscopy (fNIRS, Da Silva Ferreira Barreto et al., 2020) and applied to predicting audience preferences of different television broadcasts (Dmochowski et al., 2014) and tracking information propagation as a function of attention while listening to and reading stories (Regev et al., 2019).

Audiovisual musical performance has not been widely studied using ISC. One study investigated shared physiological responses across audience members during a live string quintet performance and explored ISC activity using musical features (Czepiel et al., 2021). Moreover, two studies yielding somewhat divergent results relate to what one might expect to find. One study with novice western observers used ISC to compare activity during audiovisual presentation of a solo dance to the conditions of audio (music) and video (dance) presentation (Jola et al., 2013). This study used novice western observers and Indian dance; a Bharatanatyam dancer accompanied by the music of Theeradha Vilaiyattu Pillai by Subramania Bharathiyar. Results showed that compared to unisensory conditions, audiovisual presentation of dance and music showed greater areas of ISC in bilateral visual and auditory cortices, with activity extending into multisensory regions in the superior temporal cortex. Regions of ISC were not found outside of occipital and temporal cortex. Another study (Abrams et al., 2013) with novice western observers used ISC to examine activity during listening to symphonic music of the late baroque period. Results showed that listening to music activated an extensive network of regions in bilateral auditory midbrain and thalamus, primary auditory and auditory association cortex and right lateralized structures in frontal and parietal cortex and motor regions of the brain. The difference in ISC found to music stimuli is striking between the two studies. While the difference in area of ISC can potentially be explained by differences in auditory properties of the music or familiarity of the participants to the genre of the music, or even particular aspects of how maps of ISC were obtained statistically, it raises the question of how much the area of ISC might vary between different musical

pieces. Another question that arises is whether the ISC found in visual cortex with dance will be found with ensemble music performance. Achieving regions of ISC has been argued to rely upon the different viewers observing the same thing at the same time (Hasson et al., 2004). In carefully edited Hollywood-style movies (Hasson et al., 2008b) it is possibly not surprising that viewers are watching the flow of events in synchrony. For dance (Jola et al., 2013) it was also argued that choreography typically directs observers' viewing. These results are distinctly different from reports of ISC based on eight participants watching an unedited video of an outdoor concert from a fixed viewpoint, where less than 5% of cortex (primarily in auditory and visual cortices) was found in the ISC maps (Hasson et al., 2010). For the case of observing a string quartet, it is not clear whether the activity of the ensemble will produce brain activity in a way that would enable regions of ISC to be found extensively in visual, or other brain regions.

In this study we use fMRI to examine the brain response to hearing and seeing an ensemble performance of a string quartet. Visual presentation of the performance is achieved through display of the musicians as stick figures derived from motion capture data. Stick figures were chosen over natural video since the stick figures allow us to control the visual salience of the players so there would not be a preference to view any particular player based on surface appearance. While being scanned participants had no task and were told simply to enjoy the performance. The data were analyzed using ISC to reveal which brain areas had correlated activity among the group of observers. They were also examined to see whether a relationship could be found between brain activity and physical measures of the sound and vision of the string quartet. Brain regions revealing ISC amongst observers indicate how brain activity becomes time-locked to external stimuli (Hasson et al., 2004), with sensory regions typically showing activity synchronized to audio (Abrams et al., 2013) and visual features (Noble et al., 2014). Examining correlation of brain activity with audio and visual signals enables us to more closely examine the influence of different features. Of particular interest is whether ISC maps would be found in visual processing areas since experiencing the string quartet by free viewing does not require a group of observers to share a common visual processing strategy. In addition, this study aims to use an audiovisual musical stimulus to investigate the underlying neural basis of a variety of auditory features. Similar to the experiment of Abrams et al. (2013), this study will use an extended piece of music to investigate these features. However, this study will also introduce vision using an upper body stick figure display of the musicians as they play their instruments. As in Alluri et al. (2012) an inter-subject correlation will be done, to investigate the ISC map resulting from the stimulus. Examination of the correlation of auditory and visual features to brain activity will also be conducted to investigate how much of the ISC can be explained by the different features. The auditory feature analysis will consist of an analysis of three different time-varying

auditory features: loudness, spectral centroid (Spectral Center of Gravity or SCG) and sound periodicity (Ap0). Loudness is primarily a measure of the psychological correlate of the amplitude of the sound wave. SCG describes the average amplitude of a sound and is connected to the brightness, or timbre of a sound. Sound periodicity is a measure of harmonics. The features investigated in the current study are based on features extracted by Giordano et al. (2013). They measured SCG, loudness and Ap0 in short naturalistic stimuli, although they refer to Ap0 as harmonicity (HNR). Visual feature analysis will consist of an analysis of speed and movement similarity, where speed is a group measure of the magnitude of the motion of the different players in the ensemble and movement similarity is the similarity of the motion of the different players.

Materials and methods

Participants

The fMRI data were collected from 18 healthy participants (eight males) with a mean age of 26 years. All subjects had normal hearing and normal or corrected to normal vision. All participants were right-handed, as assessed by the Edinburgh handedness inventory (Oldfield, 1971). Recruitment information included criteria for non-expert or novice musical listeners. Participants either received a number of course credits, or a small cash payment for their participation. Ethical approval for this study was obtained from the College of Science and Engineering ethics committee, at the University of Glasgow.

Stimulus

The stimulus was a 116 s audiovisual presentation of a string quartet; musicians were displayed as stick figures of the upper body and bow (Figure 1) in a movie presented at 25 fps. The size of the stimulus was approximately 22.8 degrees of visual angle wide and 7.4 degrees high. The width of a player was approximately 3.7 visual degrees and the distance between players was 4.6, 6.5, and 6.5 degrees going left to right, respectively, from the first violin. The visual stimulus was created from 3D motion capture of a recording of the string quartet the 'Quartetto di Cremona,' playing the allegro of String Quartet, No. 14 in D minor by Schubert. The particular piece was selected because it exemplifies a range of interaction within the quartet, including where musicians tend to play in unison, but over which the first violinist emerges progressively through a subtle original motif. Another section used a fugato style, where there is no leading part as such. Moreover, the piece is a staple of the quartet's repertoire and thus could be played consistently at a high quality. The recording has been used in previous research of Glowinski (2013) and Glowinski et al. (2014) and is described in more detail by Badino et al. (2014). Recording

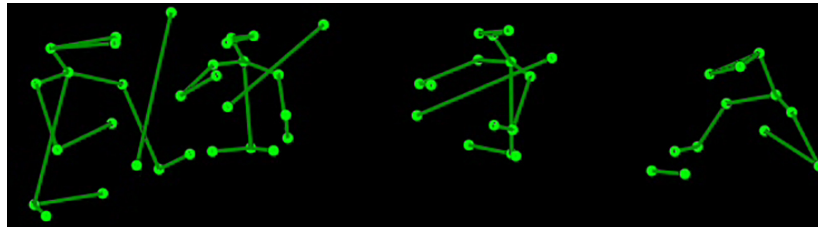


FIGURE 1
The upper body stick figure display of the string quartet the 'Quartetto di Cremona'.

was done by the InfoMus research group at the University of Genoa and took place in a 250-seat auditorium that provided an ecologically valid location for the performance. Audio and visual information were synchronized using EyesWeb XMI software (Camurri et al., 2005; Glowinski, 2013).

Auditory analysis of the string quartet

Auditory analysis was performed using custom-made Matlab scripts. Three different features were extracted from the audio of the stimulus, for the feature analysis: Loudness (measured in sones), SCG (measured in Equivalent Rectangular Bandwidth rate units) and Ap0 (measured in dB). Loudness and SCG were derived from time-varying specific loudness of the sound signals that were computed using the model of Glasberg and Moore (2002). Ap0 was computed using the YIN model of De Cheveigné and Kawahara (2002).

For each feature the median statistics were obtained for each 2 s segment of the 116 s performance. This provided 58 values for each feature that corresponded to the number of functional brain volumes obtained. However, only the first 57 values were used due to concerns of accumulated timing errors leading to a mismatch in the synchronization between the scanner volumes and the movie file that could lead to artifactual computations in the final volume. Analyses were then conducted for each of the features (Ap0, loudness and SCG) to see whether their time varying activity could account for changes in the fMRI signal. Exploratory analyses revealed little difference between the mean and the median as well as the standard deviation and interquartile range. Therefore, following (Giordano et al., 2013) we present only results based on the median.

Visual analysis of the string quartet

Visual analysis was performed using custom-made scripts in Matlab. A visual analysis of the kinematic features of the string quartet focused on analysis of the speed of the different player's motions as well as the similarity of the motion of the different players. Given the complex motion of each player there

are many possibilities of what motion feature to examine. We chose to focus on the three-dimensional motion of the bow of each player to characterize movement. The bow is the end-effector of a complex kinematic chain and has been useful to characterize emotional arm movements (Pollick et al., 2001) as well as musical conducting movements (Luck and Sloboda, 2009; Wöllner and Deconinck, 2013). To perform this, we took the original 100 Hz data sample and first calculated the average x - y - z position of the two markers on the bow for each player. We then filtered these data with a 6th order dual-pass Butterworth filter with a cutoff frequency of 10 Hz to remove high frequency noise from the data. These filtered data were then differentiated to find the three-dimensional velocity vector. From this velocity vector, we calculated the scalar value of speed, corresponding to the rate at which a bow was moving through three-dimensional space. This calculation of speed was done for each bow. Because the sampling rate for the fMRI was one sample every 2 s, we averaged the speed over 2 s epochs to find a total of 58 average speed values to represent a visual speed signal that could be compared to brain activity. We were also interested in whether the similarity of the four players might be related to brain activity, and to examine this we performed a measure of movement similarity based on a calculation of the average correlation of speed between players for each of the 58 epochs calculated. As there were four musicians there were six pairwise comparisons to make between players. Each of the six comparisons were calculated and the median was chosen as an overall measure of movement similarity. However, as with the auditory analysis, only the first 57 values were used due to concerns of accumulated timing errors leading to a mismatch in the synchronization between the scanner volumes and the movie file that could lead to artifactual computations in the final volume.

Procedure

The string quartet stimulus was presented in a single run using the software Presentation (Neurobehavioral systems, Inc.). The video of the string quartet performance was preceded and followed by 20 s of a blank screen, resulting in a run consisting of

20 s of blank screen, followed by 116 s of stimuli and ending on 20 s of blank screen. The scanning session included acquisition of fMRI data for other experiments not related to this study and this experiment came second in the run order after the first experiment of approximately 5 min. At the completion of all functional data experimental runs an anatomical scan was performed. The video was projected using an LCD projector on a translucent screen, while participants watched them in an angled mirror in the scanner. The soundtrack of the video was presented using in-ear headphones (model S14 by Sensimetrics, Malden, United States). Participants were given no explicit task when experiencing the videos, they were instructed simply to relax and enjoy the video. After scanning, participants were asked about their subjective experience of the performance and none reported familiarity with the piece.

Data acquisition and preprocessing

Data were acquired using a Siemens 3 Tesla TIM Trio magnetic resonance imaging scanner (Erlangen, Germany). Functional scans were taken to obtain brain volumes with 32 sagittal slices covering the brain volume, with $3\text{ mm} \times 3\text{ mm} \times 3\text{ mm}$ voxel size (Echo Planar 2D imaging PACE; TR = 2,000 ms; TE = 30 ms; imaging matrix of 70×70). The 116 s stimulus was contained in 58 volumes and was preceded and followed by 10 volumes for a total of 78 volumes. At the end of the functional scans, high resolution anatomical T1 weighted 3D magnetization prepared rapid acquisition gradient echo (MP- RAGE) images were acquired, with 192 sagittal slices, a 256×265 matrix size and a $1\text{ mm} \times 1\text{ mm} \times 1\text{ mm}$ voxel size and a TR of 1,900 ms.

The data files were pre-processed using Brainvoyager QX V2.8 (Brain Innovation B.V., Maastricht, Netherlands). The functional images underwent slice scan time correction, using cubic spline interpolation. 3D Motion correction was then also applied using trilinear detection and sinc interpolation. This was followed by normalization of functional scans into the common Talairach space (Talairach and Tournoux, 1988), and co-registration of functional and anatomical data. Spatial smoothing was then obtained using a Gaussian filter with a full width and half maximum (FWHM) of 6.00 mm. Finally, the functional data were trimmed to contain only the 58 volumes during the string quartet performance and these data were converted into MATLAB (The Mathworks, Natick, MA, United States) based mat files for ISC analysis.

Inter-subject correlation analysis

The ISC analysis was conducted using MATLAB and the techniques developed by Kauppi et al. (2010, 2014) and Pajula et al. (2012). As in Kauppi et al. (2010), an ISC test statistic

was derived by computing Pearson's correlation coefficient voxel-wise across the time-courses of every possible subject pair and then averaging the result across all subject pairs. A non-parametric bootstrap test with 1 million resamples was conducted to test against the null hypothesis that the test statistic was the same for unstructured data. A q -value threshold of 0.05 was used to correct for multiple comparisons using False Discovery Rate (FDR). A further cluster threshold of 108 mm^3 (four functional voxels) was applied. All 58 volumes were used in the calculation of ISC.

Relating auditory and visual features to brain activity

We used the following procedure to test whether the visual features of speed and movement similarity as well as the auditory features (Ap0, loudness, SCG) related to changes in the BOLD signal. Z- transformed data from each feature were convolved with the modeled hemodynamic response function (HRF). This gave a predictor for brain activity for each of the 57 time points considered. This model was then applied to the entire brain. Activation for the statistics of each of the features was analyzed using a random effects GLM, with Z- transform. The activations reported arise from use of an FDR adjusted criterion threshold of $q(\text{FDR}) < 0.01$.

Results

Intersubject correlation analysis

The results of the ISC analysis are shown in **Figure 2** and details of the cluster and peak voxel for this ISC map are provided in **Table 1**, along with information about the location of the peak voxel by the Talairach Daemon (Lancaster et al., 1997, 2000). A total of 8 ISC clusters were found across the 18 participants. These occurred primarily in auditory and visual sensory areas. These clusters contained a total of 154,728 voxels, or 9.9% of the total 1,562,139 anatomic voxels (1 mm^3) included in the entire brain volume.

The largest cluster had a peak in the right STG and extended bilaterally through regions in the occipital cortices, to the left STG. This cluster thus contained extensive regions of brain areas known to be involved in audio, visual and audiovisual processing. More focal clusters were also found in cortical and subcortical structures. In the cortex, clusters were found both in motor cortex (BA4) at the boundary of Brodmann Area 6 and bilateral parietal cortex. Subcortical clusters were found bilaterally both in thalamus and the basal ganglia. In the thalamus activity was found in an auditory region, the medial geniculate nucleus (MGN). In the basal ganglia activity was found in the right hemisphere to have peak ISC in the

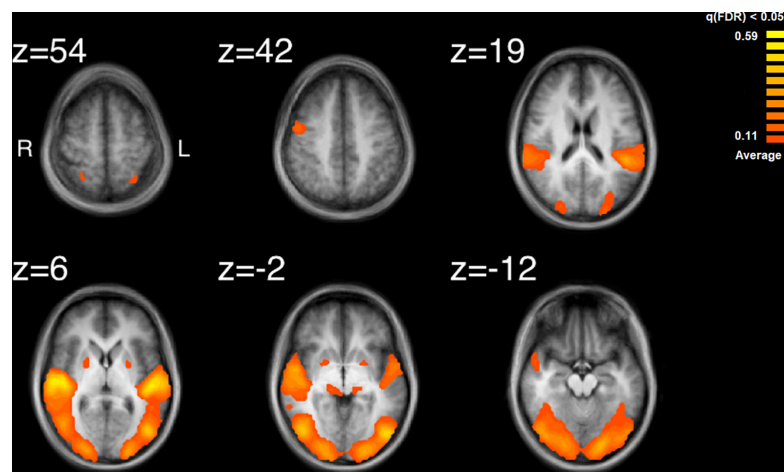


FIGURE 2

Results of the inter-subject correlation obtained from the average of all possible pairwise correlation, r , values. Significant regions of ISC include the bilateral superior temporal gyrus, peaking in the right STG, bilateral, occipital areas, the right precentral gyrus and regions in the midbrain. Images are shown in radiological format: left brain to image right and right brain to the image left.

TABLE 1 Results of the inter-subject correlation (ISC) analysis for watching and listening to the point-light string quartet.

Structure	Hemisphere	Talairach coordinate (x, y, z)	Number of voxels	Peak statistic	BA
Superior temporal gyrus	Right	(47, -14, 3)	149747	0.59	22
Precentral gyrus (PG)	Right	(50, -5, 46)	1508	0.20	4
Precuneus	Right	(27, -49, 51)	124	0.11	7
Globus pallidus	Right	(20, 0, -3)	749	0.13	
Medial geniculate	Right	(14, -26, -3)	955	0.17	
Nucleus (MGN)					
Medial geniculate	Left	(-13, -26, -3)	424	0.16	
Nucleus (MGN)					
Putamen	Left	(-22, 1, 0)	558	0.12	
Superior parietal lobule (SPL)	Left	(-31, -53, 48)	663	0.13	7

The number of voxels is the total voxel (1 mm³) count. The x, y, z coordinates are of the peak voxel in Talairach space. The peak statistic is that value in the cluster which, for each voxel, had the greatest value of average correlation. BA refers to Brodmann's Area.

globus pallidus and extended into the putamen; in the left hemisphere it had peak ISC in the putamen and appeared contained in this structure.

Auditory feature analysis

The auditory analysis investigated the median of the values of the three features, Ap0, SCG, and loudness (the trace of these features as a function of time can be seen in [Supplementary Figure 1](#)). The results for the median values of Ap0 are not reported, as no related neural activity was found for this feature. Both the features of SCG and loudness were found to predict BOLD activity. A total of 8 significant clusters were found for the median values of SCG, across both hemispheres ([Figure 3](#) and [Table 2](#)) with a total of 24,499 voxels. The most widespread activation came from the right STG with a total

voxel count of 8,004. Extensive activation was also found in the left STG, with the second highest voxel count. The right inferior temporal gyrus, cerebellum and the left inferior occipital gyrus also showed marked activations. Several large areas of activation were found for the convolution of the HRF with loudness ([Figure 4](#) and [Table 3](#)) with a total voxel count of 51,297. The largest two clusters occurred bilaterally in the STG (Brodmann areas 13, 41). The right inferior temporal gyrus along with the lentiform nucleus also contained large clusters with high voxel counts.

Visual feature analysis

The visual analysis investigated the median of the values of the two features, bow speed and a measure of the movement similarity of the players (the trace of these features as a function

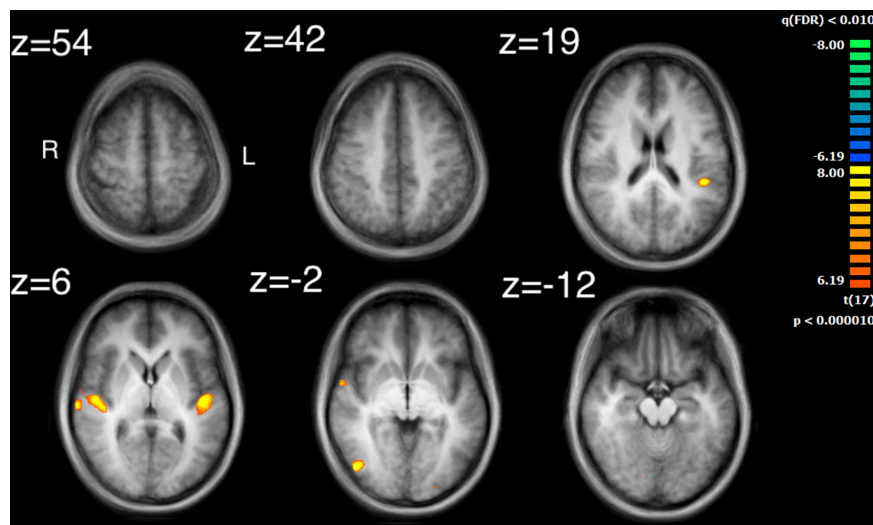


FIGURE 3

Activity for the median values of SCG. Significant regions of activity include the left and right superior temporal gyrus, left inferior occipital gyrus and right inferior temporal gyrus.

TABLE 2 Results of the auditory feature analysis for the median values of SCG.

Structure	Hemisphere	Talairach coordinate (x, y, z)	Number of voxels	Peak statistic	BA
SCG					
Superior temporal gyrus	Right	(51, -13, 1)	8004	9.601	22
Inferior temporal gyrus	Right	(42, -70, 1)	3156	10.813	19
Tuber, cerebellum	Right	(49, -58, -23)	113	5.926	
Superior frontal gyrus	Right	(21, 59, 32)	314	-6.620	9
Declive, cerebellum	Right	(15, -76, -14)	3048	7.420	
Inferior occipital gyrus	Left	(-24, -88, -2)	1301	6.783	18
Superior temporal gyrus	Left	(-42, -34, 16)	7476	10.396	41
Middle occipital gyrus	Left	(-39, -67, 4)	1087	6.789	37

The number of voxels is the total voxel (1 mm^3) count. The x, y, z coordinates are of the peak voxel in Talairach space. The peak statistic is that value in the cluster which, for each voxel, had the greatest value of average correlation. BA refers to Brodmann's Area.

of time can be seen in [Supplementary Figure 1](#)). For bow speed, results showed that the median values of bow speed predicted large bilateral clusters of activation in subcortical structures as well as bilateral activation in the posterior insula that extended into the STG ([Figure 5](#) and [Table 4](#)). The larger of the two clusters was found in the left hemisphere and the smaller was found in the right hemisphere. For the movement similarity analysis, four clusters of activation were found bilaterally, contained within the STG ([Figure 6](#) and [Table 5](#)). However, in distinction to the other features examined, movement similarity clusters were negative, meaning that BOLD activity was inversely related to movement similarity; decreasing movement similarity predicted increases in BOLD activity and vice versa. The largest clusters were found in the right hemisphere, with smaller clusters found in the left hemisphere (Brodmann 22).

Comparing intersubject correlation to auditory and visual feature analyses

Comparison of the ISC analysis ([Figure 2](#) and [Table 1](#)) and the feature analyses ([Figures 3–6](#) and [Tables 2–6](#)), show that the activation where ISC is significant encompasses several regions of activation found in the feature analyses. The STG was one of the most extensive regions of activation found in the ISC. STG clusters were found in the analysis of the median values of loudness, SCG and movement similarity. Activation found in the midbrain for the ISC only encompassed activation found for the median values of loudness and bow speed. Regions of correlated activity in the lentiform nucleus, specifically in the globus pallidus, for the ISC corresponded to significant regions of activation for the median values of SCG, loudness, and speed,

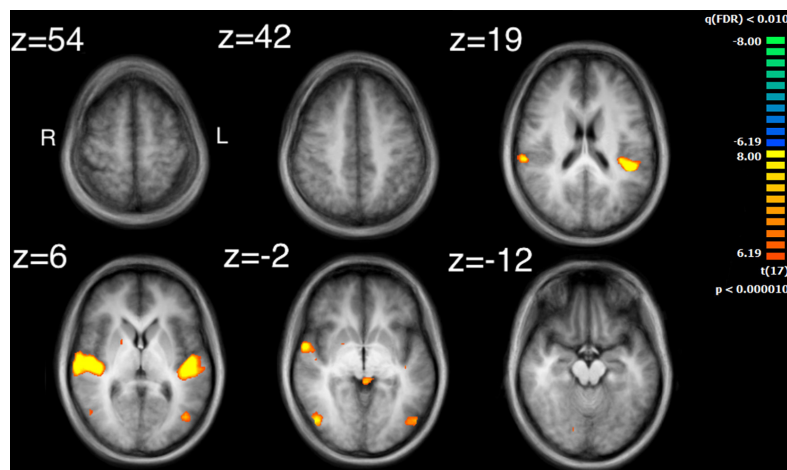


FIGURE 4

Activation for the median values of loudness. Significant regions of activity include the bilateral superior temporal gyrus, the right inferior temporal gyrus and left middle occipital gyrus.

TABLE 3 Results of the auditory feature analysis for the median values of loudness.

Structure	Hemisphere	Talairach coordinate (<i>x, y, z</i>)	Number of voxels	Peak statistic	BA
Loudness					
Superior temporal gyrus	Right	(45, -19, 7)	15982	11.554	13
Precentral gyrus	Right	(51, -1, 43)	472	5.433	6
Inferior temporal gyrus	Right	(46, -70, -2)	4184	8.633	37
Tuber, cerebellum	Right	(49, -58, -23)	921	6.809	
Declive, cerebellum	Right	(24, -58, -17)	4418	6.896	
Lentiform nucleus - lateral globus pallidus	Right	(18, 2, 1)	3805	7.502	
Lingual gyrus	Right	(15, -88, 4)	144	4.812	17
Culmen, cerebellum	Left	(-6, -31, -5)	1460	7.709	
Medial frontal gyrus	Left	(0, -4, 64)	517	5.570	6
Declive, cerebellum	Left	(-5, -85, -20)	238	5.208	
Lentiform nucleus - putamen	Left	(-21, 2, 10)	1090	6.784	
Superior temporal gyrus	Left	(-42, -34, 16)	14607	14.868	41
Middle occipital gyrus	Left	(-42, -67, 4)	3306	7.854	37
Tuber	Left	(-51, -52, -23)	153	5.143	

The number of voxels is the total voxel (1 mm³) count. The *x, y, z* coordinates are of the peak voxel in Talairach space. The peak statistic is that value in the cluster which, for each voxel, had the greatest value of average correlation. BA refers to Brodmann's Area.

although this activation was stronger for speed than for loudness or SCG.

to the lack of neural results observed for Ap0 but contrary to the observed brain activity in the left and right STG for movement similarity.

Correlation of physical features

Values for the physical auditory and visual features were correlated (Table 6). Strong correlations were observed between loudness and SCG, loudness and speed, as well as speed and SCG. Movement similarity and Ap0 both showed a lack of correlation to other visual and auditory features, corresponding

Discussion

Here we showed brain activity, revealed by fMRI, correlated among a group of observers when they experienced the audiovisual presentation of a string quartet musical performance. The present study had two aims. The first aim

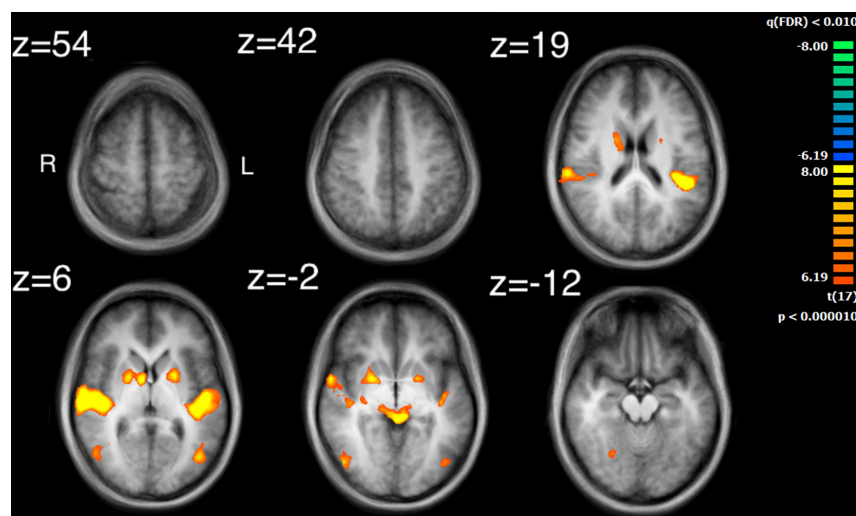


FIGURE 5

Activation for the median values of speed. Significant regions of activity include the bilateral posterior insula extending into superior temporal gyrus as well as right inferior temporal and left middle occipital gyrus in addition to numerous subcortical structures.

TABLE 4 Results of the visual feature analysis for the median values of speed.

Structure	Hemisphere	Talairach coordinate (<i>x, y, z</i>)	Number of voxels	Peak statistic	BA
Speed					
Insula	Right	(39, -19, 4)	11601	13.064	13
Inferior temporal gyrus	Right	(46, -70, 1)	1370	8.147	19
Declive	Right	(27, -61, -17)	720	7.471	
Lentiform nucleus – lateral globus pallidus	Right	(18, 2, 1)	4577	9.927	
Thalamus – pulvinar	Left	(-3, -31, -2)	2416	10.947	
Superior frontal gyrus	Right	(3, -4, 64)	514	7.174	6
Lentiform nucleus – putamen	Left	(-21, 2, 10)	2335	10.842	
Insula	Left	(-42, -37, 19)	10073	23.486	13
Middle occipital gyrus	Left	(-42, -67, 4)	1242	8.390	37

The number of voxels is the total voxel (1 mm³) count. The *x, y, z* coordinates are of the peak voxel in Talairach space. The peak statistic is that value in the cluster which, for each voxel, had the greatest value of average correlation. BA refers to Brodmann's Area.

was to investigate which regions were found in participants' ISC maps when watching an audiovisual video of a string quartet. Three auditory features (Ap0, SCG, and loudness) and two visual features (speed and movement similarity) were then extracted from the audiovisual stimulus and, subsequently, we examined how much of the ISC could be explained by the regions of activity correlated with these auditory and visual features.

Intersubject correlation maps

Intersubject correlation was found as expected in both auditory and visual sensory areas and extended into audiovisual regions in superior temporal cortex (Ethofer et al., 2006). Further regions of ISC were found in subcortical, parietal and frontal regions which implicates an extended set of brain areas

involved in the processing of complex audiovisual stimuli. Several regions of the ISC map can be attributed to processing the audio of the string quartet. For example, the medial geniculate nucleus is known for auditory processing in the thalamus (Jiang et al., 2013). In addition, there were substantial regions of ISC over bilateral primary and secondary audio areas. Indeed, the highest values of ISC found were located in the right STG (BA22), which is an important area for the processing of basic auditory features. These results are consistent with previous measurements of ISC for listening to music and watching dance with music (Abrams et al., 2013; Jola et al., 2013). Whether this auditory processing revealed by the ISC was enhanced by the visual signal is a possibility, as EEG evidence has shown that visual input can facilitate audio processing (Klucharev et al., 2003; Van Wassenhove et al., 2005; Jessen and Kotz, 2011). Similarly, Jola et al. (2013) reported more extensive ISC in auditory regions during audiovisual presentation of

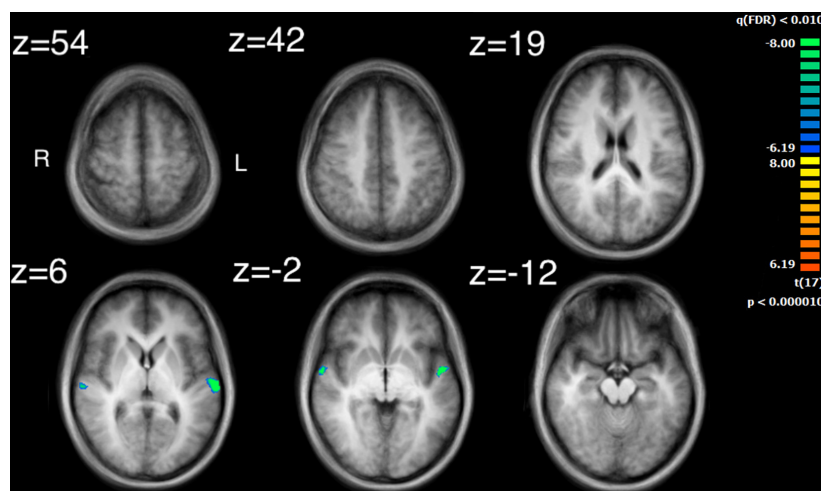


FIGURE 6
Activation for the median values of movement similarity. Regions of activity are in the left and right superior temporal gyrus. Note that clusters found for movement similarity are negatively correlated.

TABLE 5 Results of the visual feature analysis for the median values of movement similarity.

Structure	Hemisphere	Talairach coordinate (x, y, z)	Number of voxels	Peak statistic	BA
Movement similarity					
Superior temporal gyrus	Right	(60, -19, 4)	178	-8.227	22
Superior temporal gyrus	Right	(57, -4, 1)	173	-8.855	22
Superior temporal gyrus	Left	(-51, -4, 1)	441	-10.211	22
Superior temporal gyrus	Left	(-63, -19, 4)	665	-12.008	22

The number of voxels is the total voxel (1 mm³) count. The x, y, z coordinates are of the peak voxel in Talairach space. The peak statistic is that value in the cluster which, for each voxel, had the greatest value of average correlation. BA refers to Brodmann's Area.

dance with music than with music alone. The bilateral ISC found in the parietal cortex around the intraparietal sulcus has been reported previously in listening to music (Abrams et al., 2013). Similar to these earlier findings we also report a greater area of ISC in the left hemisphere. An interpretation of this parietal activity is that it reflects tracking the structure of the sound (Levitin and Menon, 2003). However, similar parietal activity has also been implicated in attentional tracking of visual objects (Culham et al., 1998).

Several regions of the ISC map can be attributed to processing the video of the string quartet. Evidence of visual processing comes from extensive ISC found bilaterally from regions of primary visual cortex through to secondary visual areas including those involved in motion processing. These findings are consistent with previous examinations of ISC for visual and audiovisual presentation of dance (Jola et al., 2013; Herbec et al., 2015; Reason et al., 2016; Pollick et al., 2018). In particular, the middle temporal region contained in the ISC map has been found to covary with the global motion of the dancer (Noble et al., 2014). This finding might help to explain how substantial correlation was found throughout the visual cortex

during free viewing of the string quartet. One consideration is that with free viewing there is no reason to believe that observers would be viewing the same musician at the same time, which is a basic requirement for visual activity to be correlated across observers. However, one possibility is that due to the large receptive fields reported for the perception of biological motion (Ikeda et al., 2005; Thompson and Baccus, 2012), extending 4 degrees from fixation, that particularly for

TABLE 6 Pearson correlations between all physical auditory and visual features.

	Ap0	SCG	Loudness	Speed	Movement similarity
Ap0	1				
SCG	0.229	1			
Loudness	-0.027	0.763***	1		
Speed	0.171	0.659***	0.879***	1	
Movement similarity	-0.071	-0.034	0.116	0.191	1

***Indicates p-value < 0.001.

the more closely spaced players there is overlap and response is being driven by a mixture of signals from multiple players. Similarly, visual neurons responsive to biological motion have been shown to integrate over long time periods (Neri et al., 1998) and thus activity might further be combined across fixations on different players. In contrast, there is evidence from eye-tracking studies that music, and sudden changes in music, can affect visual attention (Mera and Stumpf, 2014), as shown in reading (Zhang et al., 2018). Thus, another possibility is that at key events in the musical composition, musical features directed visual attention toward the same musician, though any effects are likely to be difficult to disentangle, given the documented influence of expertise, familiarity and immediate musical goal on visual processes involved in music reading in performance (Puurtinen, 2018; Vandemoortele et al., 2018; Fink et al., 2019; Bishop et al., 2021). Further studies examining eye movements could possibly help to resolve this issue.

Several regions of the ISC map can be attributed to a role for the motor system in processing the string quartet. Evidence for the involvement of motor systems come from the finding that regions of the motor cortex and basal ganglia were contained in the ISC map. The basal ganglia, including the putamen are important regions that connect to the primary motor cortex, premotor cortex and the supplementary motor area. Although ISC has previously been observed in premotor cortex and supplementary motor area when listening to classical music (Abrams et al., 2013), the finding of ISC in the basal ganglia is novel. One possible reason why ISC in the putamen was not found previously for listening is that visual stimulation is critical. However, presentation of music and dance (Jola et al., 2013) also did not reveal ISC in subcortical regions and thus visual motion coordinated with sound does not seem sufficient. Potentially a key difference between dance and a string quartet is that the dance is choreographed to the music but does not produce the music; in distinction the string quartet produced the sounds and thus there was a clear mapping between the visual and auditory stimuli. If, for example, vision of the string quartet provided a useful signal to predict the beat structure of the music then this could explain the result since the putamen has been shown to modulate its activity with prediction of the beat structure of an auditory signal (Grahn and Rowe, 2009, 2013). Similarly, greater activation has been found in the premotor cortex when listening to preferred versus non-preferred tempos (Kornysheva et al., 2010).

Auditory and visual feature analysis and their comparison to intersubject correlation maps

Investigation of regions of activity for the auditory and visual features revealed several areas, such as STG, which appeared in the ISC maps. This suggests that some of the ISC

could be explained by the auditory features contained within the music or visual features displayed by the string quartet. Median values of both SCG and loudness activated the left STG in the primary auditory cortex, although activation from loudness had a much larger cluster of activation. Median values for movement similarity showed bilateral activation of the STG, with larger clusters in the right hemisphere. ISC in the STG can probably be partly explained by the features of SCG, loudness, and movement similarity. However, the activity found in the STG by the ISC was much more widespread, whereas peak activation for loudness and SCG was centered on the primary auditory cortex. These results support the findings of Giordano et al. (2013) who also found activation for loudness in the primary auditory cortex lateralized to the left hemisphere. Alluri et al. (2012) found more extensive activation in the STG for a variety of features, including activation in area 22 of the STG, bilaterally, for the features: fullness, brightness, activity and timbral complexity. The present study also found that bilateral activation in the STG was inversely related to movement similarity. Taken together, this suggests that other features might additionally explain ISC found in the STG, as well as ISC of the STG in the right hemisphere. The findings also are consistent with the idea of hierarchical processing in auditory perception, with low level features being processed by core auditory areas (Kaas and Hackett, 2000). In addition to the STG's involvement in auditory processing, it has also been implicated as a critical structure in social cognition. Research by Ceravolo et al. (2021) revealed a widely ramified network of brain regions in decoding or inferring communicative intent in musical context, where STG took a key role even under altered perceptual conditions. Adjacent to the STG is the superior temporal sulcus, which has been implicated in visual processing and social cognition (Allison et al., 2000). Related to this observation is the fact that the brain areas related to the feature of movement similarity were found along the STG in BA22, which comprises the superior temporal sulcus. Suggesting that this movement similarity relates to social processing among the players. Overall, our results confirm and extend, using naturalistic stimuli, how this specific underlying process may also support the perception of musical ensembles.

Significant activity in the midbrain for loudness and speed could explain the ISC map regions found in the right midbrain for the ISC. ISC in the lentiform nucleus, specifically in the globus pallidus, in both the left and right hemispheres could be explained by SCG, loudness, and speed. This is a subcortical structure that seems to play a role in motor processes, although research into this area is limited and has been conducted mostly on primates and rats, with focus on neurotransmitters in nervous disorders such as Catalepsy and Parkinsonism (Wei and Chen, 2009; Galvan et al., 2010).

There was a large overlap of activity between the features SCG, loudness and speed. This may be due to the significantly high correlations found between these features. Some regions

found in the ISC maps can be linked to regions of activity found for SCG, loudness, speed and movement similarity. However, other unaccounted for activations may have occurred in response to features that were not investigated in this analysis, such as pitch, or rhythm.

Results of the present study support the findings of previous studies of other complex performances. Bilateral clusters of activation were found in the occipitotemporal area (BA19 and BA37), for median values of loudness, SCG and speed. This area has previously been implicated in human motion perception in previous research relating the visual motion of the silhouette of a dancer to brain activity (Noble et al., 2014). Jola et al. (2013) also found ISC in this area when participants watched the same dance as used by Noble et al. (2014). Grosbras et al. (2012) found that this occipitotemporal area had the greatest chance of being activated in biological motion perception experiments. Given the established role of visual motion in activating this brain region it can be questioned whether the finding of brain activity related to loudness and SCG arises purely from their correlation with visual speed. One possibility would be that the relationship between speed and auditory properties is obligatory due to physical coupling of the features. However, the study of sound production of violins indicates that SCG and speed are positively correlated but that loudness is not correlated to speed (Edgerton et al., 2014). Instead, loudness is predominately modulated by moving bow position toward the body of the instrument. A speculative explanation would be that the coupling of speed is connected to the expressive nature of the movements, which is consistent with the finding that the peak of activity related to speed was found in the posterior insula. Overall, significant open questions remain about whether or not it may be possible to identify overall neural responses to auditory features that correspond to higher-level architectonic musical structures.

Limitations

The use of the extended audiovisual string quartet stimulus in this study provides valuable information about the effects of an extended multi-modal stimulus on auditory feature analysis. It does, however, make the results of the auditory analysis less clear. The use of extended complex stimuli can be problematic, in that it is difficult to separate out which perceptual or cognitive processes might correspond to which features. It might be that a particular cortical area is found to have a correlation with a perceptual feature, but it is not necessarily the case that these two are causally related. There may well be a third or hidden variable that drives the neural activity. This can be especially unclear when the functions of cortical or subcortical areas are not well understood, or when these areas have been implicated in a variety of different processes.

Given that not all of the vast number of previously studied auditory features were explored in this study, regions of activation shown on the inter-subject correlation that did not correspond to an auditory feature in this study, may well be implicated in other auditory features not investigated here. Many other studies have placed an important emphasis on pitch, for example. These were measured in both Alluri et al. (2012) and Giordano et al. (2013). This study did not allow for a measurement of pitch, due to the nature of the auditory stimulus, a distinct pitch would have very rarely been able to be extracted, leading to confusion over which of the sound's frequencies would have determined the pitch. Rhythm and tonality were also measured by Alluri et al. (2012), but were not considered in this study.

Expertise can also have a great effect on perception. Previous studies have demonstrated the effect that expertise can have on perception of performances of dance and music. Calvo-Merino et al. (2005) for example showed that experts in one style of dance exhibited greater bilateral activity in several cortical areas related to motor perception, in comparison to controls inexperienced in dance. Gebel et al. (2013) has also recently shown this effect in musicians, finding instrument specific levels of cortical activation between pianists and trumpet players. This clearly demonstrates the effect that expertise can have on neural processing during a task. With regards to feature specific differences in neural processing, Shahin et al. (2008) found a significant enhancement of gamma band activation in musicians for the timbre of their own instrument, using EEG. The current study did not rigorously assess level of expertise in participants past indicating that they were non-expert or novice musical listeners, meaning that the subject group could have been comprised of either a group of more or less musical experience than were investigated in other studies. This may have had an effect on results obtained, as the level of expertise of participants can clearly have an effect on neural activity underlying perceptual processes. In addition to expertise, participants were not rigorously screened for familiarity with the musical piece presented, though on debrief no participant spontaneously reported familiarity with the piece. It has been shown that familiarity arising from multiple presentations of humor reduces ISC (Jääskeläinen et al., 2016). Future research could more carefully control for these factors of expertise and experience.

Obtaining a sufficient sample size to produce reliable results and satisfactory statistical power is a common challenge in fMRI group analysis. Whilst this study made use of 18 participants, in general a sample size of 20–30 is recommended for reliable results in ISC analysis using fMRI (Pajula and Tohka, 2016). Thus, although just outside the margins, the smaller sample size used in this study may affect generalizability and reproducibility of the results found.

Conclusion

We have explored the ISC of brain activity in a group of observers when watching and listening to ensemble play of a string quartet, and related these findings to the correlation of brain activity to audio and visual features of the performance. Results showed ISC in auditory and visual brain areas that could be anticipated from sensory processing. ISC was also found in parietal cortex, which has been attributed to tracking a signal over time. Finally, cortical and subcortical motor areas that have been related to perceiving rhythmic structure were also implicated. In particular, the left putamen was found both in ISC maps as well as related to the auditory feature of loudness and the visual feature of speed, suggesting that audiovisual interactions of the performance were driving this putamen activity. One possible explanation of the audiovisual interaction comes from the fact that the physical properties of speed and loudness were correlated to each other. However, for the case of movement similarity there was no correlation of this visual feature to an auditory feature. The visual property of movement similarity was related to brain activity in auditory cortex (BA22), with decreasing movement similarity among players leading to higher brain activity. Suggesting that one possible source of audiovisual interaction might be guided by the holistic property of movement similarity, which forms a salient visual feature to facilitate auditory processing (Klucharev et al., 2003; Van Wassenhove et al., 2005; Jessen and Kotz, 2011). These results raise exciting possibilities for future studies that leverage the capability of a relatively short (116 s) scan to reliably localize brain areas related to the processing of naturalistic performance.

One promising avenue for such future research would be the exploration of the ways in which musical structure may lead to convergence and divergence in musicians' actions and concomitant neural activity. Passages that are largely homophonic could be contrasted with passages of varying degrees of polyphonic complexity as well as interstitial passages that are texturally preparatory or transitional. This should help shed light on neural activity underpinning hierarchically-organized joint action involving formation and dissolution of temporary "coalitions" of musical agents.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by Ethics Committee of the College of

Science and Engineering, University of Glasgow. The patients/participants provided their written informed consent to participate in this study.

Author contributions

DG, AC, and FP originally conceived the study. AL, DN, DG, BG, and FP performed the data analysis. DG and AC originally obtained the string quartet data who generously provided them for use in this study. BG provided methods for the physical analysis of the auditory signal. IC provided expertise for interpretation of the results in a musical context. AL, DN, and DG contributed to writing the first draft. FP supervised the project. All authors contributed to editing and revising the manuscript.

Acknowledgments

We acknowledge the musicians from the Quartetto di Cremona. We thank the reviewers and the editor for their advice and encouragement to improve the manuscript. Scanning resources were provided by the Centre for Cognitive Neuroimaging, University of Glasgow.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnins.2022.921489/full#supplementary-material>

References

- Abrams, D. A., Ryali, S., Chen, T., Chordia, P., Khouzam, A., Levitin, D. J., et al. (2013). Inter-subject synchronization of brain responses during natural music listening. *Eur. J. Neurosci.* 37, 1458–1469. doi: 10.1111/ejn.12173
- Alho, K., Rinne, T., Herron, T. J., and Woods, D. L. (2014). Stimulus-dependent activations and attention-related modulations in the auditory cortex: a meta-analysis of fMRI studies. *Hear. Res.* 307, 29–41. doi: 10.1016/j.heares.2013.08.001
- Allison, T., Puce, A., and McCarthy, G. (2000). Social perception from visual cues: role of the STS region. *Trends Cogn. Sci.* 4, 267–278. doi: 10.1016/S1364-6613(00)01501-1
- Alluri, V., Toivainen, P., Jääskeläinen, I. P., Glerean, E., Sams, M., and Brattico, E. (2012). Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm. *NeuroImage*, 59, 3677–3689. doi: 10.1016/j.neuroimage.2011.11.019
- Ardizzi, M., Calbi, M., Tavaglione, S., Umiltà, M. A., and Gallese, V. (2020). Audience spontaneous entrainment during the collective enjoyment of live performances: Physiological and behavioral measurements. *Sci. Rep.* 10:3813. doi: 10.1038/s41598-020-60832-7
- Arrighi, R., Marini, F., and Burr, D. (2009). Meaningful auditory information enhances perception of visual biological motion. *J. Vision* 9, 25–25. doi: 10.1167/9.4.25
- Badino, L., D'Ausilio, A., Glowinski, D., Camurri, A., and Fadiga, L. (2014). Sensorimotor communication in professional quartets. *Neuropsychologia* 55, 98–104. doi: 10.1016/j.neuropsychologia.2013.11.012
- Bishop, L., Jensenius, A. R., and Laeng, B. (2021). Musical and bodily predictors of mental effort in string quartet music: An ecological pupillometry study of performers and listeners. *Front. Psychol.* 12:653021. doi: 10.3389/fpsyg.2021.653021
- Brooks, A., Van der Zwan, R., Billard, A., Petreska, B., Clarke, S., and Blanke, O. (2007). Auditory motion affects visual biological motion processing. *Neuropsychologia* 45, 523–530. doi: 10.1016/j.neuropsychologia.2005.12.012
- Calvo-Merino, B., Glaser, D., Grézes, J., Passingham, R., and Haggard, P. (2005). Action observation and acquired motor skills: An fMRI study with expert dancers. *Cereb. Cortex* 15, 1243–1249. doi: 10.1093/cercor/bhi007
- Camurri, A., Volpe, G., Poli, G. D., and Leman, M. (2005). Communicating expressiveness and affect in multimodal interactive systems. *IEEE Multimedia* 12, 43–53. doi: 10.1109/MMUL.2005.2
- Cantlon, J. F., and Li, R. (2013). Neural Activity during Natural Viewing of Sesame Street Statistically Predicts Test Scores in Early Childhood. *PLoS Biol.* 11:e1001462. doi: 10.1371/journal.pbio.1001462.s002
- Ceravolo, L., Schaeferlaeken, S., Frühholz, S., Glowinski, D., and Grandjean, D. (2021). Frontoparietal, Cerebellum Network Codes for Accurate Intention Prediction in Altered Perceptual Conditions. *Cerebral Cortex Commun.* 2:tgab031. doi: 10.1093/texcom/tgab031
- Chapin, H., Jantzen, K., Kelso, J. S., Steinberg, F., and Large, E. (2010). Dynamic emotional and neural responses to music depend on performance expression and listener experience. *PLoS One* 5:e13812. doi: 10.1371/journal.pone.0013812
- Culham, J. C., Brandt, S. A., Cavanagh, P., Kanwisher, N. G., Dale, A. M., and Tootell, R. B. H. (1998). Cortical fMRI Activation Produced by Attentive Tracking of Moving Targets. *J. Neurophysiol.* 80, 2657–2670. doi: 10.1152/jn.1998.80.5.2657
- Czepiel, A., Fink, L. K., Fink, L. T., Wald-Fuhrmann, M., Tröndle, M., and Merrill, J. (2021). Synchrony in the periphery: Inter-subject correlation of physiological responses during live music concerts. *Sci. Rep.* 11:22457. doi: 10.1101/2020.09.01.271650
- Da Silva Ferreira Barreto, C., Zimeo Moraes, G. A., Vanzella, P., and Sato, J. R. (2020). Combining the intersubject correlation analysis and the multivariate distance matrix regression to evaluate associations between fNIRS signals and behavioral data from ecological experiments. *Exp. Brain Res.* 238, 2399–2408. doi: 10.1007/s00221-020-05895-8
- Dahl, S., and Friberg, A. (2007). Visual Perception of Expressiveness in Musicians' Body Movements. *Music Percept.* 24, 433–454. doi: 10.1525/mp.2007.24.5.433
- De Cheveigné, A., and Kawahara, H. (2002). YIN, a fundamental frequency estimator for speech and music. *J. Acoust. Soc. Am.* 111, 1917–1930. doi: 10.1121/1.1458024
- de Gelder, B., and Solanas, M. P. (2021). A computational neuroethology perspective on body and expression perception. *Trends Cogn. Sci.* 25, 744–756. doi: 10.1016/j.tics.2021.05.010
- Dmochowski, J. P., Bezdek, M. A., Abelson, B. P., Johnson, J. S., Schumacher, E. H., and Parra, L. C. (2014). Audience preferences are predicted by temporal reliability of neural processing. *Nat. Commun.* 5:4567. doi: 10.1038/ncomms5567
- Dmochowski, J. P., Sajda, P., Dias, J., and Parra, L. C. (2012). Correlated components of ongoing EEG point to emotionally laden attention—a possible marker of engagement? *Front. Hum. Neurosci.* 6:Article112. doi: 10.3389/fnhum.2012.00112/abstract
- Edgerton, M. E., Hashim, N., and Auhagen, W. (2014). A case study of scaling multiple parameters by the violin. *Musicae Sci.* 18, 473–496. doi: 10.1177/1029864914550666
- Effenberg, A. O., and Schmitz, G. (2018). Acceleration and deceleration at constant speed: Systematic modulation of motion perception by kinematic sonification. *Ann. N. Y. Acad. Sci.* 1425, 52–69. doi: 10.1111/nyas.13693
- Ethofer, T., Pourtois, G., and Wildgruber, D. (2006). Investigating audiovisual integration of emotional signals in the human brain. *Progress Brain Res.* 156, 345–361. doi: 10.1016/S0079-6123(06)56019-4
- Fink, L. K., Lange, E. B., and Groner, R. (2019). The application of eye-tracking in music research. *J. Eye Mov. Res.* 11:10.16910/jemr.11.2.1. doi: 10.16910/jemr.11.2.1
- Fujioka, T., Trainor, L. J., Large, E. W., and Ross, B. (2012). Internalized Timing of Isochronous Sounds Is Represented in Neuromagnetic Beta Oscillations. *J. Neurosci.* 32, 1791–1802. doi: 10.1523/JNEUROSCI.4107-11.2012
- Galvan, A., Hu, X., Smith, Y., and Wichmann, T. (2010). Localization and function of GABA transporters in the globus pallidus of parkinsonian monkeys. *Exp. Neurol.* 223, 505–515. doi: 10.1016/j.expneurol.2010.01.018
- Gebel, B., Braun, C., Kaza, E., Altenmüller, E., and Lotze, M. (2013). Instrument specific brain activation in sensorimotor and auditory representation in musicians. *NeuroImage* 74, 37–44. doi: 10.1016/j.neuroimage.2013.02.021
- Giordano, B. L., McAdams, S., Zatorre, R. J., Kriegeskorte, N., and Belin, P. (2013). Abstract encoding of auditory objects in cortical activity patterns. *Cerebral Cortex* 23, 2025–2037. doi: 10.1093/cercor/bhs162
- Glasberg, B. R., and Moore, B. C. J. (2002). A model of loudness applicable to time-varying sounds. *J. Audio Eng. Soc.* 50, 331–342. doi: 10.1177/2331216516682698
- Glowinski, D. (2013). The movements made by performers in a skilled quartet: a distinctive pattern, and the function that it serves. *Front. Psychol.* 4:841. doi: 10.3389/fpsyg.2013.00841/abstract
- Glowinski, D., Riolfi, A., Shirole, K., Torres-Eliard, K., Chiorri, C., and Grandjean, D. (2014). Is he playing solo or within an ensemble? How the context, visual information, and expertise may impact upon the perception of musical expressivity. *Perception* 43, 825–828. doi: 10.1068/p7787
- Grahn, J. A., and Rowe, J. B. (2009). Feeling the Beat: Premotor and Striatal Interactions in Musicians and Nonmusicians during Beat Perception. *J. Neurosci.* 29, 7540–7548. doi: 10.1523/JNEUROSCI.2018-08.2009
- Grahn, J. A., and Rowe, J. B. (2013). Finding and Feeling the Musical Beat: Striatal Dissociations between Detection and Prediction of Regularity. *Cerebral Cortex* 23, 913–921. doi: 10.1093/cercor/bhs083
- Grosbras, M. H., Beaton, S., and Eickhoff, S. B. (2012). Brain regions involved in human movement perception: A quantitative voxel-based meta-analysis. *Hum. Brain Mapp.* 33, 431–454. doi: 10.1002/hbm.21222
- Hasson, U., Avidan, G., Gelbard, H., Vallines, I., Harel, M., Minshew, N., et al. (2009). Shared and idiosyncratic cortical activation patterns in autism revealed under continuous real-life viewing conditions. *Autism Res.* 2, 220–231. doi: 10.1002/aur.89
- Hasson, U., Furman, O., Clark, D., Dudai, Y., and Davachi, L. (2008a). Enhanced intersubject correlations during movie viewing correlate with successful episodic encoding. *Neuron* 57, 452–462. doi: 10.1016/j.neuron.2007.12.009
- Hasson, U., Landesman, O., Knappmeyer, B., Vallines, I., Rubin, N., and Heeger, D. J. (2008b). Neurocinematics: The neuroscience of film. *Projections* 2, 1–26. doi: 10.3167/proj.2008.020102
- Hasson, U., Yang, E., Vallines, I., Heeger, D. J., and Rubin, N. (2008c). A hierarchy of temporal receptive windows in human cortex. *J. Neurosci.* 28, 2539–2550. doi: 10.1523/JNEUROSCI.5487-07.2008
- Hasson, U., Ghazanfar, A. A., Galantucci, B., Garrod, S., and Keysers, C. (2012). Brain-to-brain coupling: a mechanism for creating and sharing a social world. *Trends Cogn. Sci.* 16, 114–121. doi: 10.1016/j.tics.2011.12.007

- Hasson, U., Malach, R., and Heeger, D. J. (2010). Reliability of cortical activity during natural stimulation. *Trends Cogn. Sci.* 14, 40–48. doi: 10.1016/j.tics.2009.10.011
- Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., and Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science* 303, 1634–1640. doi: 10.1126/science.1089506
- Herbec, A., Kauppi, J. P., Jola, C., Tohka, J., and Pollick, F. E. (2015). Differences in fMRI intersubject correlation while viewing unedited and edited videos of dance performance. *Cortex* 71, 341–348. doi: 10.1016/j.cortex.2015.06.026
- Hou, Y., Song, B., Hu, Y., Pan, Y., and Hu, Y. (2020). The averaged inter-brain coherence between the audience and a violinist predicts the popularity of violin performance. *NeuroImage* 211:116655. doi: 10.1016/j.neuroimage.2020.116655
- Ikeda, H., Blake, R., and Watanabe, K. (2005). Eccentric perception of biological motion is unscalarmy poor. *Vision Res.* 45, 1935–1943. doi: 10.1016/j.visres.2005.02.001
- Iversen, J. R., Repp, B. H., and Patel, A. D. (2009). Top-Down Control of Rhythm Perception Modulates Early Auditory Responses. *Ann. N. Y. Acad. Sci.* 1169, 58–73. doi: 10.1111/j.1749-6632.2009.04579.x
- Jääskeläinen, I. P., Koskentalo, K., Balk, M. H., Autti, T., Kauramäki, J., Pomren, C., et al. (2008). Inter-subject synchronization of prefrontal cortex hemodynamic activity during natural viewing. *Open Neuroimaging J.* 2, 14–19. doi: 10.2174/187444000802010014
- Jääskeläinen, I. P., Pajula, J., Tohka, J., Lee, H. J., Kuo, W. J., and Lin, F. H. (2016). Brain hemodynamic activity during viewing and re-viewing of comedy movies explained by experienced humor. *Sci. Rep.* 6, 27741. doi: 10.1038/srep27741
- Jessen, S., and Kotz, S. A. (2011). The temporal dynamics of processing emotions from vocal, facial, and bodily expressions. *NeuroImage* 58, 665–674. doi: 10.1016/j.neuroimage.2011.06.035
- Jiang, F., Stecker, G. C., and Fine, I. (2013). Functional localization of the auditory thalamus in individual human subjects. *NeuroImage* 78, 295–304. doi: 10.1016/j.neuroimage.2013.04.035
- Jola, C., McAleer, P., Grosbras, M.-H., Love, S. A., Morison, G., and Pollick, F. E. (2013). Uni- and multisensory brain areas are synchronised across spectators when watching unedited dance recordings. *Iperception* 4, 265–284. doi: 10.1068/i0536
- Kaas, J. H., and Hackett, T. A. (2000). Subdivisions of auditory cortex and processing streams in primates. *Proc. Natl. Acad. Sci. U.S.A.* 97, 11793–11799. doi: 10.1073/pnas.97.22.11793
- Kauppi, J.-P., Jääskeläinen, I. P., Sams, M., and Tohka, J. (2010). Inter-subject correlation of brain hemodynamic responses during watching a movie: localization in space and frequency. *Front. Neuroinformatics* 4:5. doi: 10.3389/fninf.2014.00002
- Kauppi, J.-P., Pajula, J., and Tohka, J. (2014). A versatile software package for inter-subject correlation based analyses of fMRI. *Front. Neuroinformatics* 8:2.
- Keller, P. E., and Appel, M. (2010). Individual Differences, Auditory Imagery, and the Coordination of Body Movements and Sounds in Musical Ensembles. *Music Percept* 28, 27–46. doi: 10.1525/mp.2010.28.1.27
- Klucharev, V., Mötönen, R., and Sams, M. (2003). Electrophysiological indicators of phonetic and non-phonetic multisensory interactions during audiovisual speech perception. *Cogn. Brain Res.* 18, 65–75. doi: 10.1016/j.cogbrainres.2003.09.004
- Kornysheva, K., Cramon von, D. Y., Jacobsen, T., and Schubotz, R. I. (2010). Tuning-in to the beat: Aesthetic appreciation of musical rhythms correlates with a premotor activity boost. *Hum. Brain Mapp.* 31, 48–64. doi: 10.1002/hbm.20844
- Lancaster, J. L., Rainey, L. H., Summerlin, J. L., Freitas, C. S., Fox, P. T., Evans, A. C., et al. (1997). Automated labeling of the human brain: A preliminary report on the development and evaluation of a forward-transform method. *Hum. Brain Mapp.* 5, 238–242. doi: 10.1002/(SICI)1097-0193(1997)5:4
- Lancaster, J. L., Woldorff, M. G., Parsons, L. M., Liotti, M., Freitas, C. S., Rainey, L., et al. (2000). Automated Talairach atlas labels for functional brain mapping. *Hum. Brain Mapp.* 10, 120–131. doi: 10.1002/1097-0193(200007)10:3
- Langers, D. R., van Dijk, P., Schoenmaker, E. S., and Backes, W. H. (2007). fMRI activation in relation to sound intensity and loudness. *NeuroImage* 35, 709–718. doi: 10.1016/j.neuroimage.2006.12.013
- Levitin, D. J., and Menon, V. (2003). Musical structure is processed in “language” areas of the brain: a possible role for Brodmann Area 47 in temporal coherence. *NeuroImage* 20, 2142–2152. doi: 10.1016/j.neuroimage.2003.08.016
- Luck, G., and Sloboda, J. A. (2009). Spatio-temporal cues for visually mediated synchronization. *Music Percept.* 26, 465–473. doi: 10.1525/mp.2009.26.5.465
- Mera, M., and Stumpf, S. (2014). Eye-tracking film music. *Music Mov. Image* 7, 3–23. doi: 10.5406/musimoviimag.7.3.0003
- Neri, P., Morrone, M. C., and Burr, D. C. (1998). Seeing biological motion. *Nature* 395, 894–896. doi: 10.1038/27661
- Noble, K., Glowinski, D., Murphy, H., Jola, C., McAleer, P., Darshane, N., et al. (2014). Event Segmentation and Biological Motion Perception in Watching Dance. *Art Percept.* 2, 59–74. doi: 10.1163/22134913-00002011
- Nozaradan, S., Peretz, I., Missal, M., and Mouraux, A. (2011). Tagging the neuronal entrainment to beat and meter. *J. Neurosci.* 31, 10234–10240. doi: 10.1523/JNEUROSCI.0411-11.2011
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia* 9, 97–113. doi: 10.1016/0028-3932(71)90067-4
- Pajula, J., Kauppi, J.-P., and Tohka, J. (2012). Inter-Subject Correlation in fMRI: Method Validation against Stimulus-Model Based Analysis. (B. J. Harrison, Ed.). *PLoS One* 8:e41196. doi: 10.1371/journal.pone.0041196.s012
- Pajula, J., and Tohka, J. (2016). How many is enough? Effect of sample size in inter-subject correlation analysis of fMRI. *Comput. Intell. Neurosci.* 2016, 2094601. doi: 10.1155/2016/2094601
- Palmer, C. (2012). “Music performance: Movement and coordination,” in *The Psychology of Music*, 3rd Edn, ed. D. Deutsch (Amsterdam: Academic Press), 405–422. doi: 10.1016/B978-0-12-381460-9.00010-9
- Petrini, K., McAleer, P., Neary, C., Gillard, J., and Pollick, F. E. (2014). Experience in judging intent to harm modulates parahippocampal activity: An fMRI study with experienced CCTV operators. *Cortex* 57, 74–91. doi: 10.1016/j.cortex.2014.02.026
- Petrini, K., Pollick, F. E., Dahl, S., McAleer, P., McKay, L., Rocchesso, D., et al. (2011). Action expertise reduces brain activity for audiovisual matching actions: An fMRI study with expert drummers. *NeuroImage* 56, 1480–1492. doi: 10.1016/j.neuroimage.2011.03.009
- Phillips-Silver, J., and Keller, P. E. (2012). Searching for roots of entrainment and joint action in early musical interactions. *Front. Hum. Neurosci.* 6:26. doi: 10.3389/fnhum.2012.00026
- Platz, F., and Kopiez, R. (2012). When the eye listens: A meta-analysis of how audio-visual presentation enhances the appreciation of music performance. *Music Percept.* 30, 71–83. doi: 10.1525/mp.2012.30.1.71
- Pollock, F. E., Paterson, H. M., Bruderlin, A., and Sanford, A. J. (2001). Perceiving affect from arm movement. *Cognition* 82:B51–B61. doi: 10.1016/S0010-0277(01)00147-0
- Pollock, F. E., Vicary, S., Noble, K., Kim, N., Jang, S., and Stevens, C. J. (2018). Exploring collective experience in watching dance through intersubject correlation and functional connectivity of fMRI brain activity. *Progress Brain Res.* 237, 373–397. doi: 10.1016/bs.pbr.2018.03.016
- Puurtinen, M. (2018). Eye on Music Reading: A Methodological Review of Studies from 1994 to 2017. *J. Eye Mov. Res.* 11:10.16910/jemr.11.2.2. doi: 10.16910/jemr.11.2.2
- Reason, M., Jola, C., Kay, R., Reynolds, D., Kauppi, J. P., Grosbras, M. H., et al. (2016). Spectators’ aesthetic experience of sound and movement in dance performance: A transdisciplinary investigation. *Psychol. Aesthetics Creativity Arts* 10:42. doi: 10.1037/a0040032
- Regev, M., Simony, E., Lee, K., Tan, K. M., Chen, J., and Hasson, U. (2019). Propagation of information along the cortical hierarchy as a function of attention while reading and listening to stories. *Cerebral Cortex* 29, 4017–4034. doi: 10.1093/cercor/bhy282
- Röhl, M., Kollmeier, B., and Uppenkamp, S. (2011). Spectral loudness summation takes place in the primary auditory cortex. *Hum. Brain Mapp.* 32, 1483–1496. doi: 10.1002/hbm.21123
- Röhl, M., and Uppenkamp, S. (2012). Neural coding of sound intensity and loudness in the human auditory system. *J. Assoc. Res. Otolaryngol.* 13, 369–379. doi: 10.1007/s10162-012-0315-6
- Rowland, S. C., Hartley, D. E., and Wiggins, I. M. (2018). Listening in naturalistic scenes: What can functional near-infrared spectroscopy and Intersubject correlation analysis tell us about the underlying brain activity? *Trends Hear.* 22:233121651880411. doi: 10.1177/2331216518804116
- Schachner, A., Brady, T. F., Pepperberg, I. M., and Hauser, M. D. (2009). Spontaneous Motor Entrainment to Music in Multiple Vocal Mimicking Species. *Curr. Biol.* 19, 831–836. doi: 10.1016/j.cub.2009.03.061
- Schurz, M., Radua, J., Tholen, M. G., Malisic, L., Margulies, D. S., Mars, R. B., et al. (2021). Toward a hierarchical model of social cognition: A neuroimaging meta-analysis and integrative review of empathy and theory of mind. *Psychol. Bull.* 147:293. doi: 10.1037/bul0000303

- Schutz, M. (2008). Seeing music? What musicians need to know about vision. *Empirical Musical. Rev.* 3, 83–108. doi: 10.18061/1811/34098
- Shahin, A. J., Roberts, L. E., Chau, W., Trainor, L. J., and Miller, L. M. (2008). Music training leads to the development of timbre-specific gamma band activity. *NeuroImage* 41, 113–122. doi: 10.1016/j.neuroimage.2008.01.067
- Talairach, J., and Tournoux, P. (1988). *Co-Planar Stereotaxic Atlas of the Human Brain: 3-D Proportional System: An Approach to Cerebral Imaging*. New York, NY: Thieme, doi: 10.1017/S0022215100111879
- Thompson, J. C., and Baccus, W. (2012). Form and motion make independent contributions to the response to biological motion in occipitotemporal cortex. *NeuroImage* 59, 625–634. doi: 10.1016/j.neuroimage.2011.07.051
- Van Wassenhove, V., Grant, K. W., and Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proc. Natl. Acad. Sci. U.S.A.* 102, 1181–1186. doi: 10.1073/pnas.0408949102
- Vandemoortele, S., Feyaerts, K., Reybrouck, M., De Bièvre, G., Brône, G., and De Baets, T. (2018). Gazing at the partner in musical trios: a mobile eye-tracking study. *J. Eye Mov. Res.* 11:10.16910/jemr.11.2.6. doi: 10.16910/jemr.11.2.6
- Vines, B. W., Krumhansl, C. L., Wanderley, M. M., Dalca, I. M., and Levitin, D. J. (2011). Music to my eyes: Cross-modal interactions in the perception of emotions in musical performance. *Cognition* 118, 157–170. doi: 10.1016/j.cognition.2010.11.010
- Wanderley, M. M., Vines, B. W., Middleton, N., McKay, C., and Hatch, W. (2005). The Musical Significance of Clarinetists' Ancillary Gestures: An Exploration of the Field. *J. New Music Res.* 34, 97–113. doi: 10.1080/09298210500124208
- Wei, L., and Chen, L. (2009). Effects of 5-HT in globus pallidus on haloperidol-induced catalepsy in rats. *Neuroscience letters* 454, 49–52. doi: 10.1016/j.neulet.2009.02.053
- Wing, A. M., Endo, S., Bradbury, A., and Vorberg, D. (2014). Optimal feedback correction in string quartet synchronization. *J. R. Soc. Int.* 11, 20131125–20131125. doi: 10.3758/BF03214280
- Wöllner, C., and Deconinck, F. J. (2013). Gender recognition depends on type of movement and motor skill. Analyzing and perceiving biological motion in musical and nonmusical tasks. *Acta psychol.* 143, 79–87. doi: 10.1016/j.actpsy.2013.02.012
- Zhang, H., Miller, K., Cleveland, R., and Cortina, K. (2018). How listening to music affects reading: Evidence from eye tracking. *J. Exp. Psychol.* 44, 1778–1791. doi: 10.1037/xlm0000544