



HAL
open science

Uni- and multisensory brain areas are synchronised across spectators when watching unedited dance recordings

Corinne Jola, Phil Mcaleer, Marie-Hélène Grosbras, Scott Love, Gordon
Morison, Frank E. Pollick

► **To cite this version:**

Corinne Jola, Phil Mcaleer, Marie-Hélène Grosbras, Scott Love, Gordon Morison, et al.. Uni- and multisensory brain areas are synchronised across spectators when watching unedited dance recordings. *I-Perception*, 2013, 4 (4), pp.265-284. 10.1068/i0536 . hal-02647342

HAL Id: hal-02647342

<https://hal.inrae.fr/hal-02647342>

Submitted on 29 May 2020

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.



Distributed under a Creative Commons Attribution 4.0 International License

Uni- and multisensory brain areas are synchronised across spectators when watching unedited dance recordings

Corinne Jola

INSERM-CEA Cognitive Neuroimaging Unit, NeuroSpin Center, F-91191 Gif-sur-Yvette, France, and School of Psychology, University of Glasgow, Glasgow G12 8QB, UK; e-mail: Corinne.JOLA@cea.fr

Phil McAleer

Institute of Neuroscience and Psychology, University of Glasgow, Glasgow G12 8QB, UK; e-mail: Philip.McAleer@glasgow.ac.uk

Marie-Hélène Grosbras

Institute of Neuroscience and Psychology, University of Glasgow, Glasgow G12 8QB, UK; e-mail: Marie.Grosbras@glasgow.ac.uk

Scott A. Love

Department of Psychological and Brain Sciences, Indiana University, Bloomington, Indiana, USA; e-mail: sclove@indiana.edu

Gordon Morison

Computer, Communication and Interactive Systems, Glasgow Caledonian University, Glasgow G4 0BA, UK; e-mail: gordon.morison@gcu.ac.uk

Frank E. Pollick

School of Psychology, University of Glasgow, Glasgow G12 8QB, UK; e-mail: Frank.Pollick@glasgow.ac.uk

Received 23 April 2012, in revised form 20 February 2013; published online 3 June 2013.

Abstract. The superior temporal sulcus (STS) and gyrus (STG) are commonly identified to be functionally relevant for multisensory integration of audiovisual (AV) stimuli. However, most neuroimaging studies on AV integration used stimuli of short duration in explicit evaluative tasks. Importantly though, many of our AV experiences are of a long duration and ambiguous. It is unclear if the enhanced activity in audio, visual, and AV brain areas would also be synchronised over time across subjects when they are exposed to such multisensory stimuli. We used intersubject correlation to investigate which brain areas are synchronised across novices for uni- and multisensory versions of a 6-min 26-s recording of an unfamiliar, unedited Indian dance recording (Bharatanatyam). In Bharatanatyam, music and dance are choreographed together in a highly intermodal-dependent manner. Activity in the middle and posterior STG was significantly correlated between subjects and showed also significant enhancement for AV integration when the functional magnetic resonance signals were contrasted against each other using a general linear model conjunction analysis. These results extend previous studies by showing an intermediate step of synchronisation for novices: while there was a consensus across subjects' brain activity in areas relevant for unisensory processing and AV integration of related audio and visual stimuli, we found no evidence for synchronisation of higher level cognitive processes, suggesting these were idiosyncratic.

Keywords: intersubject correlation, superior temporal gyrus, audiovisual integration, dance, novice spectators, perception.

1 Introduction

1.1 Audiovisual integration

Day to day, we are exposed to a continuous stream of multisensory audio and visual stimulation. For optimal social interaction, our brain integrates these sensory signals from different modalities into a coherent one. For example, others' movements, gestures, and emotional expressions are combined with auditory signals, such as their spoken words, to create a meaningful perception. A good illustration for such a cross-modal integration of audiovisual signals (AV) is the McGurk effect (McGurk & MacDonald, 1976), where the resulting percept is a novel creation of the visual and auditory informa-

tion available. Most cases of AV integration are less spectacular. In those cases, the use of information from multiple sensory modalities simply enhances perceptual sensitivity, allowing more accurate judgments of experts and novices on particular parameters of the sensory stimulus (e.g. Arrighi, Marini, & Burr, 2009; Jola, Davis, & Haggard, 2011; Love, Pollick, & Petrini, 2012; Navarra & Soto-Faraco, 2005).

The area in the human brain that has predominantly been identified as the loci of AV integration by means of functional magnetic resonance (fMRI) is the superior temporal sulcus (STS; Beauchamp, Argall, Bodurka, Duyn, & Martin, 2004; Kreifelts, Ethofer, Grodd, Erb, & Wildgruber, 2007; Kreifelts, Ethofer, Huberle, Grodd, & Wildgruber, 2010; Szycik, Tausche, & Münte, 2008), in particular in its posterior (Möttönen et al., 2006) and ventral parts of the left hemisphere (Calvert, Campbell, & Brammer, 2000), sometimes extending into the posterior superior temporal gyrus (STG). Further indication of the crucial role STS plays in AV integration comes from magnetoencephalography (Raij, Uutela, & Hari, 2000), PET (Sekiyama, Kanno, Miura, & Sugita, 2003), and ERP (Reale et al., 2007) studies. Other brain areas showing enhanced activity in AV conditions include the medial temporal gyrus (MTG; Kilian-Hütten, Vroomen, & Formisano, 2011; Li et al., 2011), the insula, the intra parietal sulcus (IPS; see Calvert, 2001), and the pre-central cortex (Benoit, Raij, Lin, Jääskeläinen, & Stufflebeam, 2010). Moreover, many authors agree that most sensory brain areas are multimodal (e.g. Klemen & Chambers, 2011), whereby multimodal describes the responsiveness to more than one sensory modality. However, using fMRI, the requirements to evidence AV integration in multimodal brain areas are more specific (Calvert, 2001) but also disputed (e.g. Love, Pollick, & Latinus, 2011).

In fMRI, brain activity is localised by task-related increase or decrease in blood-oxygen-level dependence (BOLD). Enhanced activity in response to coherent AV stimulation above the sum of the activation by the unisensory A and V stimuli is one of the criteria for AV integration (i.e. superadditivity). As the initial AV integration principles were based on single neuron animal studies, they do not fully apply to fMRI (Klemen & Chambers, 2011). For instance, superadditivity in fMRI can potentially be missed if individual neurons within a voxel are not integrative and it might thus not be an appropriate means of verification (e.g. Beauchamp, 2005; Love et al., 2011), as specifically shown by Wright, Pelphrey, Allison, McKeown, and McCarthy (2003) for STS. Furthermore, superadditivity can sometimes be found because of reduced activity to unisensory stimuli: Laurienti et al. (2002) showed overadditive responses for AV stimulation that were based on a deactivation of the auditory and visual cortex by cross-modal visual and auditory stimulation. Hence, despite extensive research in AV integration, the functional processes of multisensory brain areas are still debated. Moreover, the diversity of results may not only lay in the different analyses methods (Beauchamp, 2005; Love et al., 2011) but also in the different experimental designs and stimuli used (see Calvert, 2001).

To identify brain areas where AV integration takes place, most designs involved explicit evaluative behavioural tasks. For instance, in Kreifelts et al. (2007), subjects had to classify the emotion of people speaking single words based on visually (facial expression) and/or auditory (affective speech prosody) information. This approach allows direct association of brain activity with behavioural evidence of AV integration, such as higher classification hit rates or faster reaction times. In everyday life, however, we do not continuously evaluate others' expressions into emotional categories yet multisensory stimuli are integrated nevertheless, and largely independent of attentional resources (see Kreifelts et al., 2010). Thus, findings from multisensory integration under passive viewing conditions are a better approximation of mechanisms present in real life and therefore fewer assumptions made when generalising the results.

Furthermore, most AV studies focused on short stimuli predominantly consisting of faces and voices while everyday experiences consist of integrated continuous AV information from various objects and body parts over extended periods of time. In order to validate and better understand the functional processes of previously identified multisensory integration areas, it is thus important to use more natural complex multidimensional AV stimuli. Hence, in contrast to classical neuroimaging studies on AV integration, the stimuli should support implicit processing (not rely upon subjects' behavioural responses), be of relatively long duration, and recorded during so-called "natural" viewing. Natural viewing in this context refers to free viewing (i.e. with no predefined fixation points) of complex scenes of moving stimuli with a longer duration that is closer to real life than precisely parameterised stimuli. Intersubject correlation (ISC) is one method alongside other recent developments in neuroscience such as independent component analysis (Bartels & Zeki, 2004; Wolf, Dziobek, & Heekeren, 2010), Wavelet correlation (Lessa et al., 2011), or event boundary analysis (Zacks et al.,

2001; Zacks, Speer, Swallow, & Maley, 2010), which enable analyses of fMRI data recorded while presenting stimuli that fulfil these criteria (Hasson, Nir, Levy, Fuhrmann, & Malach, 2004).

1.2 Intersubject correlation

ISC is a measure of how similar subjects' brain activity is over time. In their seminal study, Hasson et al. (2004) found activity in well-known visual and auditory cortices as well as in high-order association areas (STS, lateral temporal sulcus, retrosplenial and cingulate cortices) to be correlated when subjects watched a 30-min segment of the film "The Good, the Bad and the Ugly." Some of the areas that Hasson et al. (2004) identified have not previously been associated with sensory processing of external stimuli. Many subsequent ISC studies further supported the finding that the extent of the correlation in the identified brain areas was indeed determined by the external stimulus that the subjects were exposed to. For example, the areas of ISC were found to be more extended for structured, edited feature films than for realistic one-shot, unedited recordings of an everyday life scene (Hasson et al., 2008b). In fact, to explore cortical coherence for AV stimuli within and between different groups of audiences and film genres, many of the following ISC studies employed edited feature films (Furman, Dorfman, Hasson, Davachi, & Dudai, 2007; Hasson, Furman, Clark, Dudai, & Davachi, 2008a; Hasson, Vallines, Heeger, & Rubin, 2008c, 2009a; Hasson et al., 2004, 2008b; Kauppi, Jääskeläinen, Sams, & Tohka, 2010). Based on the differences found in ISC for specific films, it was even suggested that ISC could be a measure for the films' effectiveness to drive collective audience engagement.

As well as accessing the effects that specific films have on spectators, there are several scientific reasons to invest in novel approaches such as ISC. First, because of the nature of our environment: long complex moving stimuli are closer to everyday multisensory experiences. Analyses based on the general linear model (GLM) typically require short and repeated presentation of the stimuli in order to conform to the model-based approach. GLM also treats activity in areas that are not task related but consistently activated across subjects as error variance and may not identify these as regions of interest (Hejnar, Kiehl, & Calhoun, 2007). Model-based approaches are thus problematic for the analysis of brain responses to natural viewing of complex long stimuli with no a priori assumptions (Bartels & Zeki, 2004). Second, because of the nature of the human brain: measuring functional coherence across spectators by means of data-driven approaches like the ISC is independent of the level of brain activity (Hejnar et al., 2007) and does not rely on estimation of haemodynamic response functions (Jääskeläinen et al., 2008). This is relevant since the haemodynamic response function varies between individuals as well as between brain areas (Aguirre, Zarahn, & D'Esposito, 1998; Handwerker, Ollinger, & D'Esposito, 2004). Hence, ISC is a highly reliable, selective and time-locked activity from natural viewing and allows studying cortical activity without prior assumption on their functionality (Hasson, Malach, & Heeger, 2009b). Therefore, ISC is a particularly effective method to investigate AV processing as it circumvents a number of issues outlined above: task, stimulus duration, and assumptions on criteria of integration.

1.3 Choreographed but unedited dance stimuli

The choice of our stimuli was driven by our goal to extend previous research on ISC in a multisensory integration context. Our first aim was to compare the correlation of BOLD activity between subjects for A, V, and AV stimulation when the stimuli come from the same source. Dance is naturally viewed with as well as without music, and thus an optimal stimulus to investigate brain activity in response to these modalities. Second, we aimed at using unedited recordings, to prevent ISC based on cuts and close-ups. This was possible by choosing an established dance choreography. In dance, the means of transmitting a story and directing the spectators' attention is the choreography, the way in which movements are set and creatively combined with the music. This also gives a dance piece its dramatic structure within which sound and actions are related. Notably, this is very different from sounds and actions that naturally occur together, such as the "clang" of a hammer hitting a nail; but the relationship between the sound and the movement in the dance is not random. Lastly, our final goal was to investigate up to which level of uni- and multisensory processing the BOLD responses are correlated between subjects for unfamiliar stimuli. We therefore chose an Indian dance form, Bharatanatyam, which tells a story by gestural dance movements. In Bharatanatyam, the music and movements are interdependent (Vatsyayan, 1963) but unfamiliar to most spectators. While Hindu-specific emotional expressions, as used in Bharatanatyam, have been found to be universally understood in isolation (Hejmadi, Davidson, & Paul, 2000), here, they are embedded in a novel compositional structure. Furthermore, as Pillai (2002) states, it requires an extensive knowledge to fully comprehend the

story: “Bharatanatyam is highly codified and requires a significant level of connoisseurship in order to be understood beyond a surface level.” Hence, we chose to expose our subjects to both uni- and multisensory versions of an unedited but choreographed performance of a classical Indian dance piece in the style of Bharatanatyam. This particular performance was unfamiliar to naïve spectators and its narrative undecipherable (see Reason & Reynolds, 2010). It allowed us to measure brain areas in which subjects process an unfamiliar dance presented in an unedited recording in a coherent manner.

This approach to study multisensory integration using ISC is novel. No study has yet experimentally investigated the BOLD time-course specific to AV integration. Hasson et al. (2008b) compared ISC of a silent film (V) with the ISC of a different audio-book soundtrack (A). The authors found overlapping multisensory areas including STS, temporal parietal junction, and IPS in the left hemisphere. Importantly, however, the two sensory stimuli came from different sources and thus give no indication of ISC in AV integration. In another study, the authors measured the ISC of subjects looking at an unedited recording of a public space and found much less activity than for an edited film in the primary visual, auditory and lateral occipital areas than for an edited feature film (Hasson et al., 2008b). However, the unedited film was a recording of an everyday scene that had no crafted dramatic structure. We predicted that the choreographed dance would show ISC in visual and auditory unisensory, multisensory, and AV integration areas despite its unedited format. Finally, as we used a dance form that was unfamiliar to the subjects, we did not expect higher order cognitive and/or motor areas to significantly correlate. This prediction was based on the principle of equivalent brain functions in a group of people who are exposed to the same stimuli, up to the level at which consensus on the meaning of the stimuli is given. While the movements and music of the Indian dance are available to all subjects on a sensory level, the associated meaning of the combined sensory information varies between novice spectators (Reason & Reynolds, 2010). Hence, a similar processing between subjects can be expected on a sensory level, whereas brain activity at higher levels of cognition is anticipated to be idiosyncratic and thus less correlated. Importantly though, sound and musical gestures were found to jointly enhance an audience’s reception of a piece during music performance (Vines, Krumhansl, Wanderley, & Levitin, 2006; Vines, Krumhansl, Wanderley, Dalca, & Levitin, 2011), similar to enhanced performance for AV stimuli. We thus expected that when dance is accompanied by music, the audience is more likely to understand the narrative of a performance. For example, if people understand a speaker better by integrating his or her verbal expressions with his or her gestural actions, they come to more similar conclusions of what the person is saying based on multisensory integration—and the more coherent a performance is perceived to be, the more neuronal activity is expected to correlate between spectators. We thus expected more correlation for AV than for A or V.

Theoretically, increased coherence can be postulated at the low level of unisensory processing (A and V), at the level of multisensory integration (AV; Grosbras, Beaton, & Eickhoff, 2011) or at the level of action understanding (fronto-parietal; Grafton, 2009; McNamara et al., 2008). We thus correlated the activity across the whole brain but expected significant correlation between subjects in auditory areas when listening to the music (see Koelsch, 2011), in visual areas when watching the dance (see Grill-Spector & Malach, 2004), and in AV areas (i.e. the posterior STS) when watching the dance performance accompanied by music. We also explored the correlation across modalities, i.e. whether the audio and visual stimuli evoked similar responses when presented alone. For instance, the group average of the A to V correlation of each individual subject would reveal whether basic sensory features of music and movement shared a common structure so that exposure to one form of sensory stimulation, either music or movement, would yield a similar BOLD response as being exposed to the other. Finally, a limitation of ISC is that it may contain significant correlation between subjects based on coherent responses to artefacts or noise. It is thus relevant to compare the areas with those that have previously had been identified in uni- and multisensory processing. For this, we applied a GLM subtraction and conjunction analysis (see Friston, 2005) on our own data of long segments to compare our results. Although this required ignoring some GLM model restrictions, it allowed a more direct comparison of brain areas that show enhanced activity (as identified by means of GLM) with areas that significantly correlate between subjects (as identified by means of ISC).

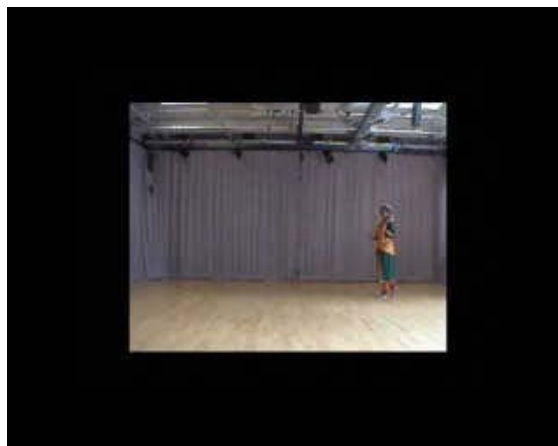
2 Materials and methods

2.1 Subjects

Twelve naïve observers (between 18 and 25 years, all but one from the UK, 50% males) passively watched a video of a woman performing a Bharatanatyam solo (classical Indian dance). All subjects were right handed, had normal or corrected-to-normal vision, and received payment for their participation. The subjects had no hearing problems, no musical training, and were not familiar with Indian dance. The study was approved by the Ethics Committee of the Faculty of Information and Mathematical Sciences, University of Glasgow. All subjects gave their written informed consent prior to inclusion in the study.

2.2 Stimuli

The stimulus material was derived from a standard definition (720×576 pixels, 25 fps) recording of a solo Bharatanatyam dance performed in appropriate costumes by a semi-professional Indian dancer of 6 min and 26 s (see AV example in [Movie 1](#) or at <http://pacoweb.psy.gla.ac.uk/watchingdance>, and [Figure 1](#) for an illustration of all three conditions).



Movie 1. AV version.

Bharatanatyam has its roots in south India, in the state of Tamil Nadu. A traditional Bharatanatyam performance lasts about two hours and consists of seven or more sections. We used a “padam” section which is widely regarded as the most lyrical, involving aspects of love such as the love of a mother for a child. In our example, the story was of a woman’s struggle with an especially naughty boy. It is told by use of hand gestures, connotative facial expressions, and spatial shifts in the vertical axis and different directions of body movement in space accompanied by typical music that also involves singing in Tamil (e.g. Pillai, [2002](#)). None of our participants understood the language. The *padam* is also known to be primarily a musical composition, where the dancer represents the music by synchronising the gestures to the melody. The dancer mimes the lyrics using the eyebrows, eyelids, nose, lips, chin and coordinates the movements of the head, chest, waist, hips, and feet with the musical notes in a highly complex manner (Vatsyayan, [1963](#)).

The music used here was Theeradha Vilayattu Pillai by Subramanya Bharathiyar from Nupura Naadam. It entails vocal lyrics, rhythmic use of cymbals (Nattuvangam) attached to the dancer’s feet as well as a drum (Mrudangam), a violin (held and played differently than in Western classical music), and a flute. This song is popularly used for Bharatanatyam and is based on a four-beat rhythm cycle (“tala”) where each line of lyric fits into four beats with increasing speed from 52 to 62 bpm, thus deflecting the slower and more regular beats of the scanner background noise. The score was according to an original composition by Subramania Bharati, a revolutionary Tamil poet in the early 20th century. The composer is acknowledged to have had a huge impact on Carnatic music. In contrast to Western music, there is no absolute pitch in the Carnatic system. The melody (“raga”) is based on five or more musical notes. Importantly, how the notes are played to create a particular mood is more important in defining a raga than the notes themselves.

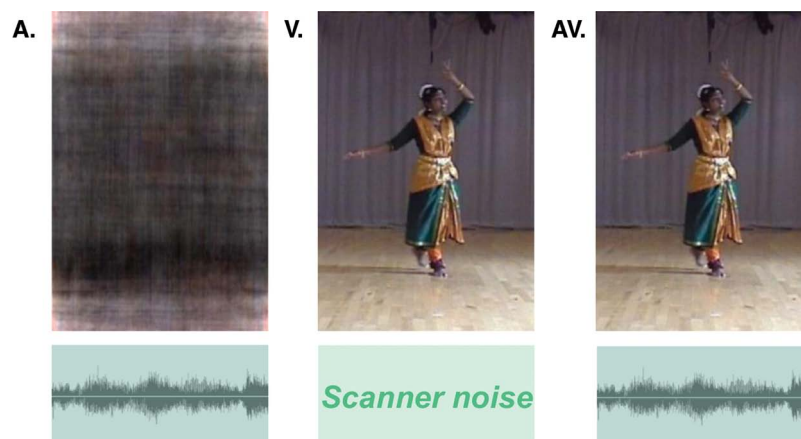


Figure 1. Sensory conditions audio (A), visual only (V), and audiovisual (AV). To balance audio and visual sensory stimulations over all three conditions, subjects saw a still uninformative picture created from the video in the audio condition (i.e., corresponding luminance structure but randomised phase for the chrominance components). The background scanner noise is present in all three conditions but predominantly in those conditions without music.

2.3 Conditions

Subjects were presented the complete duration of the dance three times in randomised order: once with both the soundtrack and visual dance, once with just the visual dance, and once with just the soundtrack. During the soundtrack-only condition, a static image that matched the luminance and spatial frequency content of a typical video frame was displayed on screen (Figure 1). Subjects were simply asked to enjoy the performance in all three conditions while their eye movements were monitored by the experimenter to make sure they watched the display throughout the experiment. The video was exported to a 540×432 pixels video that covered a field of view of $20^\circ \times 17^\circ$ of the fMRI compatible NNL (Nordic NeuroLab) goggles. The audio was presented via NNL headphones with an average of 75 dB.

2.4 Scanning procedure

Each subject had two scanning sessions of approximately 1 hour separated by a short break. Session 1 was dedicated to another free-viewing experiment and included the acquisition of a high-resolution T1-weighted anatomical scan using a 3D magnetisation prepared rapid acquisition gradient recalled echo (MP-RAGE) T1-weighted sequence (192 slices; 1-mm cube isovoxel; Sagittal Slice; TR = 1900 ms; TE = 2.52; 256×256 image resolution). Session 2 consisted of the functional run where T2*-weighted MRIs were acquired continuously (EPI, TR = 2000 ms; TE = 30 ms; 32 slices; 3-mm cube voxel; FOV = 210 mm, 70×70 image resolution) using a 3T Siemens Magnetom TIM Trio scanner with acoustic noise reduced by up to 20 dB compared with other systems and a 12-channel Siemens head coil. Subjects were wearing NNL headphones (www.nordicneurolab.com), which have a further 30 dB passive noise reduction. The functional run was in total 600 volumes and included the three dance videos (193 volumes per presentation) with a 10-s period of black screen with a central white fixation cross between conditions and 10 s at the beginning of the run and 12 s at the end. A short anatomical scan was also included in this session.

2.5 Analysis

A standard pipeline of pre-processing of the functional data was performed for each subject (Goebel, Esposito, & Formisano, 2006). For both pre-processing and analysis, we used BrainVoyager QX (Version 2.1, Brain Innovation B.V., Maastricht, the Netherlands). Slice scan time correction was performed using sinc interpolation. In addition, 3D motion correction was performed to detect and correct for small head movements by spatially aligning all the volumes of a subject to the first volume using rigid-body transformations. Estimated translation and rotation parameters never exceeded 3 mm, or 3 degrees, except for one female subject who was excluded from the data analysis due to excessive head motion during scanning (>4 mm). Finally, the functional MR images were high-pass temporal filtered with a cut-off of four cycles (i.e. 0.0033 Hz) and smoothed spatially using a Gaussian filter (FWHM = 6 mm). The first five volumes of the functional scans were excluded to elimi-

nate any potential effects of filtering artefacts. The data were then aligned with the AC–PC (anterior commissure–posterior commissure plane) and transformed into Talairach standard space (Talairach & Tournoux, 1988). To transform the functional data into Talairach space, the functional time-series data of each subject were first co-registered with the subject's 3D anatomical data of the same run and then co-registered with the anatomical from the first run, which had been transformed into Talairach space. This step results in normalised 4D volume time-course (VTC) data. Normalisation was performed combining a functional–anatomical affine transformation matrix, a rigid-body AC–PC transformation matrix, and an affine Talairach grid scaling step. The functional data were analysed using the “model-free” approach for ISC within and across conditions, and with the GLM RFX approach using conjunction analysis.

For the ISC, we computed a voxel-by-voxel correlation map for each subject in all three conditions separately. For this, we chose a more conservative approach than the pairwise correlation (see Kauppi et al., 2010) by modelling the time course of each voxel of a subject with the average time course of the homologue voxel of the remaining subjects in a linear regression analysis for the entire duration of the dance, i.e. sensory condition of 193 volumes. In other words, the ISC used the activity of other subjects' time courses as the regressors and no other factors were tested, as would be in a conventional GLM with a contrast design (see Friston, 2005). The 11 resulting maps were averaged across subjects using a random-effect analysis. We repeated this analysis across conditions for the cross-modal correlations: AV to V (the time course of one subject in condition AV modelled by the average time course from condition V); AV to A (the time course of one subject in condition AV modelled by the average time course from condition A); and A to V.

For the random-effects GLM analysis on the whole run of 600 volumes, each condition was modelled as a boxcar function of 386 s, convolved with a haemodynamic response function. We used three contrast models. Parameter estimates of the AV condition were separately contrasted with A ($AV > A$) and V ($AV > V$) to assess regions preferentially responsive to visual and audio processing in a multisensory condition. To identify multisensory regions, we computed the conjunction between the two contrasts $(AV > V) \cap (AV > A)$. The conjunction analysis used here tests for the logical AND (conjunction null hypothesis; see Ethofer, Pourtois, & Wildgruber, 2006; Friston, Penny, & Glaser, 2005). AV stimuli are processed by means of their unisensory audio and visual components as well as by multisensory integration. The contrast $AV > V$ removes the activity related to unisensory visual processing in AV and reveals audio and multisensory activity. The contrast $AV > A$ eliminates the activity related to the unisensory audio processing in AV and reveals visual and multisensory activity. The conjunction analysis of $(AV > V) \cap (AV > A)$ thus subtracts unisensory processes, exposing multimodal activity; in particular if the activity is in an area not present in the unisensory conditions (see also Szameitat, Schubert, & Müller, 2011). Each significance map (at least $p < 0.001$) was corrected for multiple comparisons using a cluster-size threshold at $\alpha 0.05$ (Forman et al., 1995; Goebel et al., 2006).

3 Results

3.1 ISC within conditions

We observed significant ISC in the occipital and temporal cortices during free viewing of 6-min 26-s Indian dance, dependent on the sensory conditions present in the stimulus. As can be seen in [Figure 2](#) and [Table 1](#), the clusters of significant correlation were located in expected areas: in the audio condition, subjects' time courses of BOLD activity correlated significantly in the STG (bilateral), with the clusters centred in Heschl's gyrus. In the visual condition, subjects' time courses of BOLD activity were significantly correlated in the lingual gyrus (bilateral), the right middle occipital gyrus (MOG), the right fusiform gyrus (FFG), and the left cuneus. In the multisensory condition, we found significant correlations of subjects' time courses in the STG (bilateral), occipital MOG (bilateral), the left lingual gyrus and the right cuneus. These AV correlations overlapped with areas identified in the audio- and visual-only conditions and AV could, in fact, be viewed as a summary of both unisensory conditions, however, with some alterations. Notably, compared with A, AV showed a greater extent of the activity in STG going into its posterior parts. Compared with V, AV showed a reduction in the extension of significant correlation in the right MOG but a bilaterally greater extension in the extrastriate visual cortex consisting of the lingual gyrus, FFA, and the cuneus. Notably, for both V and AV, the areas were more extended in the left hemisphere (LH) than the right (RH).

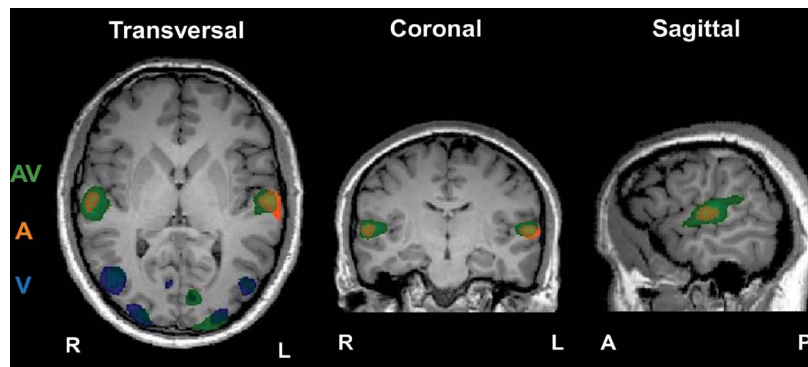


Figure 2. Areas of significant correlation between subjects within conditions A (in orange), V (in blue), and AV (in green). $p < 0.0001$ for all clusters, cluster-level threshold corrected at $\alpha 0.05$. Slices taken at Talairach coordinates $X = 59$, $Y = -15$, and $Z = 4$ for the sagittal, coronal, and transversal views, respectively.

Table 1. Clusters of significant ISC in uni- and multisensory conditions.

BA	Area	H.	Coordinates			No. voxels	T
ISC within condition A							
22	Superior temporal gyrus	R	56	-11	6	875	4.40
		L	-61	-14	6	1,747	5.29
ISC within condition V							
18	Lingual gyrus	R	8	-71	0	374	3.97
		L	-13	-80	-3	270	4.09
37	Middle occipital gyrus	R	44	-71	3	8,850	6.43
19	Fusiform gyrus	R	23	-53	-6	343	4.15
18	Cuneus	L	-25	-95	1	3,969	5.45
ISC within condition AV							
22	Superior temporal gyrus	R	59	-17	9	5,133	6.36
		L	-58	-14	6	5,864	6.32
18	Lingual gyrus	L	-13	-80	-3	6,027	5.16
37	Middle occipital gyrus	R	44	-68	3	2,156	5.47
		L	-46	-71	3	787	4.06
18	Cuneus	R	26	-92	0	1,457	4.57

Note. Coordinates = peak coordinates; R = right; L = left; BA = Brodmann area; H. = hemisphere; No. voxels = cluster size in number of voxels for all $p < 0.001$ (α uncorrected), with a cluster extent of 50 voxels. Significance levels are given in T-values (T), all $p < 0.01$, cluster-threshold corrected at $\alpha 0.05$. BA and area labelling was based on the automated Talairach Daemon system (Lancaster et al., 2000).

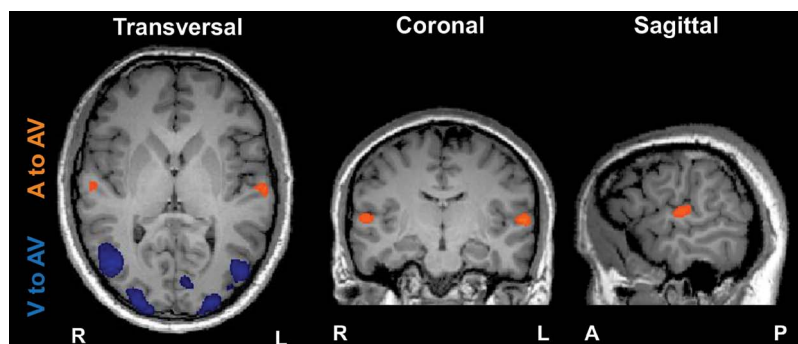


Figure 3. Areas of significant correlation between subjects across conditions A to AV (in orange) and V to AV (in blue), $p < 0.0001$ for all clusters, cluster-level threshold corrected at $\alpha 0.05$. Slices taken at Talairach coordinates $X = 59$, $Y = -15$, and $Z = 4$ for the sagittal, coronal, and transversal views, respectively.

Table 2. Clusters of significant ISC across uni- and multisensory conditions. See the note to [Table 1](#) for columns' information.

BA	Area	H.	Coordinates			No. voxels	T
ISC across conditions A to AV							
22	Superior temporal gyrus	R	59	-14	6	571	4.36
		L	-61	-14	6	878	4.77
ISC across condition V to AV							
37	Middle occipital gyrus	R	41	-68	0	13,421	5.71
18	Cuneus	L	-25	-95	3	7,617	5.44
18	Lingual gyrus	L	-13	-80	-3	1,542	4.33
37	Fusiform gyrus	R	23	-50	-9	1,031	4.17

3.2 ISC across conditions

The BOLD time courses across the sensory conditions A and AV showed significant correlation in bilateral primary auditory areas while across V and AV showed significant correlation in secondary visual areas (see [Figure 3](#) and [Table 2](#)). Compared with ISC within conditions (Section 3.1), the extent of the significant correlation was reduced in STG, but enhanced in MOG and extrastriate cortex. Notably, V to AV showed correlation only in the left lingual gyrus, consistent with the left lateralisation for this region in ISC AV. Hence, activity in primary and secondary sensory areas in response to a unisensory stimulus was partly synchronised with the activity in the same areas produced in response to the multisensory stimulus. This indicates that the processing of the multisensory stimuli at least partly conserved the separate processing characteristics of the unisensory stimuli. No areas were significantly synchronised across the two unisensory conditions (A to V).

3.3 GLM-based contrast and conjunction analyses

The GLM-based analyses revealed a pattern of results that was consistent with original expectations and that was supported by the results given in the ISC analysis (see [Table 3](#) and [Figure 4](#)). The contrast $AV > V$ represents the involvement of the auditory aspects in AV processing and showed significant enhanced bilateral activity in the STG, including the primary auditory cortex. The contrast $AV > A$ represents the connection of the visual mode within the multisensory condition and showed significantly enhanced activity in the right (superior temporal and MOG) and left (MTG, IOG, and cuneus) hemispheres. It is important to note that these contrasts are not revealing brain areas sensitive

Table 3. Significant brain activation in GLM analysis for uni- and multisensory contrasts. See the note to [Table 1](#) for columns' information.

BA	Area	H.	Coordinates			No. voxels	T
GLM contrast $AV > V$							
22	Superior temporal gyrus	R	50	-8	0	2,783	7.87
		L	-58	-35	12	1,507	7.50
Contrast GLM $AV > A$							
22	Superior temporal gyrus	R	56	-38	12	193	5.79
37	Middle temporal gyrus	L	-46	-62	9	764	8.76
37	Middle occipital gyrus	R	44	-71	3	1,419	6.60
19	Inferior occipital gyrus	L	-40	-74	-3	325	5.31
17	Cuneus	L	-10	-95	3	322	-6.20
Conjunction GLM Analysis $(AV > V) \cap (AV > A)$							
22	Superior temporal gyrus	R	56	-32	15	356	4.82
		L	-58	-38	15	138	4.34

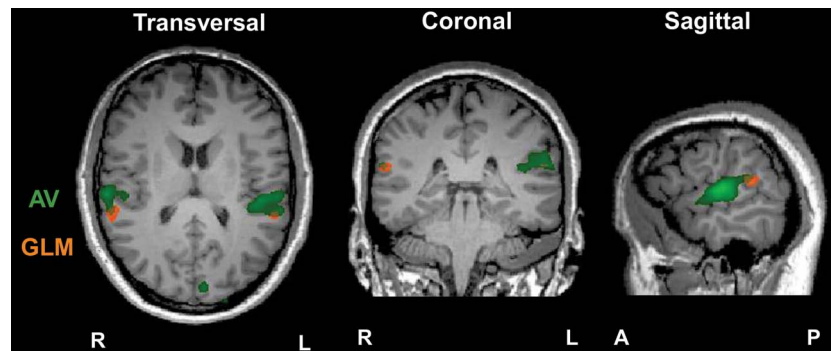


Figure 4. Significant clusters for ISC AV (green) and multisensory conjunction analysis (orange), $p < 0.001$, α level uncorrected. Slices taken at Talairach coordinates $X = 59$, $Y = -34$, and $Z = 14$ for the sagittal, coronal, and transversal views, respectively.

to the unisensory conditions alone; hence, both contrasts manifest aspects of AV: the contrast $AV > A$ also showed significant activity in areas outside the primary visual cortex (e.g. STG) and the contrast $AV > V$ showed enhanced activity over a wider area than was correlated for A. To find regions associated with multisensory processing, we computed a conjunction analysis $(AV > V) \cap (AV > A)$. This analysis revealed bilateral activity in the posterior STS (pSTS) with the significant area in the LH being further posterior than the RH.

3.4 Sensory control analysis

To verify that the areas of significant ISC were due to sensory processing rather than randomly correlated activity (e.g. resting state), we conducted an additional control analysis. For this, for each individual subject, the different sensory conditions were randomly assigned to three control groups (R1, R2, and R3). We then calculated the number of voxels synchronised for each control group. The percentage of synchronised voxels of the whole brain was then compared with the percentage of synchronised areas in AV. As visible in [Figure 5](#), the groups containing random assignment of the sensory conditions do not lead to more than 2% of synchronisation across the whole brain.

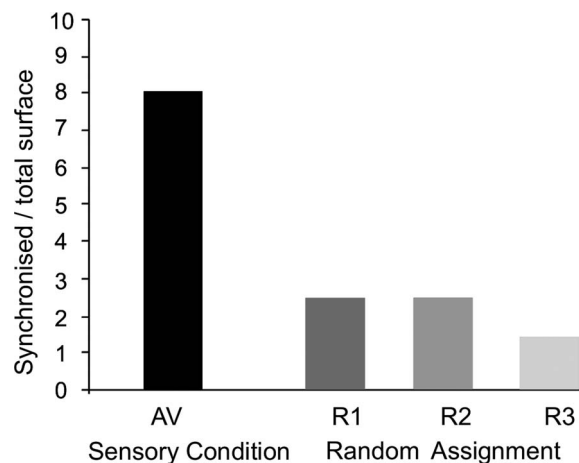


Figure 5. Percentage of total brain surface significantly correlated at $p < 0.001$, α level uncorrected for AV and three different random assignments of each functional scan to three groups (R1, R2, and R3).

4 Discussion

We examined which brain areas are significantly correlated between 11 healthy subjects in uni- and multisensory processing by means of ISC. We used free viewing of a long segment of a dance performance with and without music (AV and V) and an audio condition (A) in which subjects listened to the music only while looking at a static, uninformative picture. The dance recording did not involve any changes in the visual scene or in the rhythm of presentation created by zooming or cuts. There were thus no visual cinematic effects that directed spectators' perception. We used Bharatanatyam

dance since the narrative and compositional elements are completely unfamiliar and unknown to novices (Reason & Reynolds, 2010; Vatsyayan, 1963) even though the emotional expressions have been found to be universally understood (Hejmadi et al., 2000). A lack of familiarity does not, however, necessarily imply a complete lack of processing the composition. For example, non-signers show competence in understanding the structure of a narrative given by sign language without grasping the meaning (Fenlon, Denmark, Campbell, & Woll, 2007). This is relevant as we studied the level to which sensory processing is correlated between subjects' brain activity when recognisable gestures are embedded in a novel but structured context that is presented in an unedited format. Unlike an edited feature film, we did not expect higher order areas to be correlated; and unlike a recording of a real-life situation (Hasson et al., 2008b), dance is choreographed for movement, sound, and importantly, its combination, and we thus expected enhanced correlation in AV processing across spectators even when presented with an unedited recording.

4.1 Synchronisation and multisensory processing

Brain activity was significantly correlated across subjects in several functionally relevant regions for auditory (e.g. Heschl's gyrus), visual (e.g. lingual gyrus, MOG, cuneus), and multisensory processing (e.g. pSTG). Importantly, subjects' correlation expanded into STG—the area that is frequently reported for processing of AV conditions (Calvert, 2001; Ethofer et al., 2006). Furthermore, the multisensory area pSTS showed enhanced activity as revealed by a GLM conjunction analysis and was partly overlapping with the area that was significantly correlated between subjects in the AV condition. Conjunction analysis as a tool to study sensory processing in the human brain has been discussed widely (e.g. Ethofer et al., 2006; Szameitat et al., 2011). It is nevertheless exceptional that the GLM showed similar results to the ISC because we applied GLM in an unconventional block design: a single block for each condition lasting 6 min and 26 s. The GLM does not capture haemodynamic adaptation processes (see Ou, Raij, Lin, Golland, & Hämäläinen, 2009) and thus, normally, repeated presentations of stimuli with short durations are used in order to maximise its power. In light of both ISC and GLM conjunction analyses, our results show that multisensory processing may not only be identified by enhanced activity but extended synchronisation.

Notably, we did not find significant correlation beyond primary and secondary sensory areas. Our study differs from previous work using ISC on two particular aspects that may explain variation in the results. First, we used a non-edited video of a choreographed dance and, second, we used a more conservative ISC analysis by measuring subject-to-average correlations between several participants. In regard to the former, it is relevant to note that most studies using ISC (e.g. Bartels & Zeki, 2004; Hasson et al., 2004) investigated subjects' brain response to watching narrative movie sequences that were edited in a particular way, promoting a narrative that maximises the attention of the observer. For instance, Hasson et al. (2004) presented 30 min of the movie “The Good, the Bad, and the Ugly” and reported correlations between subjects over large regions of the brain including higher order areas. Importantly, this film is popular and well liked, using highly stylised editing with numerous scene changes and close-ups that draw the spectator into the storyline. Our study explored subjects' brain responses to long sequences of unedited stimuli which we expected to differ from results obtained from viewing edited films.

4.2 The role of STG in sensory processing

As in numerous other studies (e.g. Hein & Knight, 2008), we found the STG to be a relevant site for sensory processing. The ISC revealed significant correlation between subjects in large clusters in STG for watching dance with music (AV) and in a smaller cluster for listening to music only (A). The correlation was bilateral but more extended in the LH for audio alone, for audio as preserved in AV (ISC A to AV), as well as for AV stimuli. The GLM contrasts related to audio processing also showed bilateral enhanced activity in STG, however, predominantly in the RH. Moreover, the contrast for visual contributions (AV > A) showed unilateral enhanced STG activity in the RH only (see also Meyer, Greenlee, & Wuerger, 2011). Thus, similar to the STS in Beauchamp et al. (2004), we found hemispheric differences in STG. Importantly, the Indian dance performance consisted of gestural movements and the music involved a singer reciting a text in Tamil. One could therefore argue that the left STG activity found here was evoked by the voice in the auditory stimuli while the right STG correlation was more specifically activated in response to the visual perception of the gestures. Though for novice spectators unfamiliar with Tamil, the text was incomprehensible and it is thus unlikely that the left STG was primarily driven by verbal understanding. However, since Indian dance gestures were

found to be universally recognised, we suggest that the perception of these gestures may have modulated STG activation bilaterally; according to Calvert et al. (1997) and MacSweeney et al. (2004), who found that silent gestures activate the STG. Furthermore, Möttönen et al. (2006) found left posterior STG/STS activity during the observation of sign gestures. The activity was, however, dependent on whether the gestures have been recognised as speech. Using a novel connectivity approach, Nath and Beauchamp (2011) found evidence for dynamically modified functional changes from STG and visual cortex to STS depending on the most reliable sensory modalities. Further studies are needed which can dissociate primary from reciprocal synchronised activity to verify the role of visual and auditory stimulation in STG.

Clearly, a component of STG activity is driven by audio elements. However, the more posterior parts of STG did not show significant correlation between subjects in the audio-only condition. Importantly, the control analyses supported that correlation in pSTS is based on AV correlation, which is unlikely to be observed on the basis of unrelated audio and/or visual stimulation. First, a negligible amount of 2% across the entire brain was correlated randomly. Second, A and V stimulation was present in all conditions (e.g. scanner noise in V, visual control in A), but we found clear differences between correlation for these unisensory conditions and the AV correlation. It is thus very unlikely that the AV correlation in pSTS can be evoked by random, unrelated auditory or visual streams. Finally, many previous publications on AV integration on language and body action processing found that the pSTS activity was enhanced (e.g. Meyer et al., 2011), also when using edited movies (e.g. Wolf et al., 2010). According to Hasson et al. (2008c), who found that the STS was indicative of a coherent progressive narrative by contrasting forward to backward played films, we propose that the correlation in STS found here was modulated by the dance narrative, despite its novelty.

While significant correlation between subjects in low-level auditory and visual areas has been shown before (e.g. Hasson et al., 2004, 2008c; Lerner, Honey, Silbert, & Hasson, 2011), our study is novel by using a more systematic approach presenting A, V, and AV of the very same dance performance, showing evidence for a significant correlation of the activity in a multisensory integration area (pSTS) across spectators who watched an unedited recording. This has not been reported before. Hence, ISC seems to be functionally sensitive and has the potential to tackle a number of issues present in AV research when unedited but choreographed complex stimuli are used in their original form. To further investigate why the non-overlapping parts of pSTS showed enhanced activity in a task-related manner (GLM) but were not significantly correlated (ISC), additional studies are needed that better fulfil the criteria of GLM analysis.

The correlation between subjects in pSTS is indeed interesting considering both continuous audio and visual streams were unfamiliar. Although dance and music in Bharatanatyam are interwoven in such a way that the two arts become one coherent whole (Vatsyayan, 1963), this is not sufficient for the novice spectator to either fully comprehend the narrative in such a manner that they would correlate in higher order cognitive areas (see Section 4.4) or to perceive a common cross-modal structure (see Section 4.3). Nevertheless, subjects' multisensory integration was coherent. In other words, subjects can have a common level of AV integration that leads to idiosyncratic cognitive interpretations. Future studies testing ISC for different levels of disruptions would shed further light on the functional role of STS in the perception of dance structure.

4.3 Other synchronised areas

While the GLM conjunction analysis showed no significant enhanced activity other than in bilateral pSTS, further visual areas correlated significantly between subjects for AV stimuli. These were in the left lingual gyrus, the MOG bilateral, and the right cuneus, areas known for higher order processing of visual information. Some of these were related to visual aspects; however, the correlation was much extended in AV and, indeed, the left lingual gyrus has repeatedly been found to be activated for AV integration as well in combination with tactile perception (see Calvert, 2001). The cuneus has been shown to be involved in visual processes and participates during the switching of attention across visual features (Le, Pardo, & Hu, 1998) whereas the area in the left MOG is next to the fusiform face area (FFA), often described as occipital face area and sensitive for face processing (Gauthier et al., 2000). For visual unisensory stimuli (V, V to AV), the right FFA was also correlated. The area was medial to those parts of the FFA that have previously been identified to be activated in response to emotionally strong body postures (e.g. de Gelder, Snyder, Greve, Gerard, & Hadjikhani, 2004). In addition, part of the synchronised activity was in the vicinity of the right extrastriate body area (EBA)

that is known to be involved in the perception of human form (Downing & Peelen, 2011). In the current study using movements accompanied by music, this area was bilaterally enhanced (GLM AV > A) as well as correlated (ISC AV).

Other laterality differences between correlations for AV and for V were found in the cuneus (right for AV, left for V). Though meta-analyses of previously published work on biological motion perception did reveal lateralisation for the RH (Grosbras et al., 2011), it is relevant to note that labelling cortical structures based on group data is not unproblematic, especially for more extensive activations. We used Tailarach Daemon, an automated coordinate-based system (Lancaster et al., 2000) to label the location of the peak activation. However, the cuneus is part of the medial/inferior occipital gyrus as is the lingual gyrus, which is ventral to the cuneus on the lower bank of the calcarine sulcus. One could thus argue that within the scope of a more extended peak activity, that for both, AV and V, the middle occipital and/or the lingual gyri are bilaterally correlated in AV and V. The laterality may thus be an artefact of locating the activity rather than a representation of the actual processes. Hence, the ISC of AV stimuli showed clearly areas that are involved in visual sensory processing but integrating sound with vision led to small but notable changes in the activity across subjects of these visual areas.

Notably, though, the BOLD activity in response to our stimuli did not correlate from sound to vision. Such cross-modal sensory processing has been reported previously but in particular for visual and auditory stimuli that were associated in a more straightforward manner. For instance, visual observation of a light flash evoked an internally generated rhythm (Grahn, Henry, & McAuley, 2011). Furthermore, Bidet-Caulet, Voisin, Bertrand, and Fonlupt (2005) found activity in the temporal biological motion area when subjects were listening to a walking human suggesting hierarchical components in processing multisensory stimuli from V5 to posterior STG/STS (see also Beauchamp et al., 2004; Wright et al., 2003). However, our stimuli were much more complex. Since Krumhansl and Schenck (1997) found that music and dance share some common structural patterns, we suggest that our subjects did not link these visual and auditory properties due to the novelty and complexity of the stimuli. It is possible that mentally generated images to music as well as mentally evoked music to visual stimulations were present but uncorrelated (i.e. idiosyncratic) across spectators, suggesting that cross-modal synchronisation may only be present for stimuli with low-level correspondences and not for highly complex stimulus material.

Nevertheless, despite the fact that all of our spectators were unfamiliar with the narrative of the Indian dance, a number of unisensory areas and areas of AV integration were correlated. Interestingly though, on the one end, subjects' BOLD response in the lowest level of processing, the primary visual areas, was enhanced (AV > A) but sparsely correlated. The lack of extensive correlation in V1 could be due to the free-viewing situation where the focal point can be individual for each subject at each moment in time. Furthermore, on the other end, the activity in higher order areas was neither significantly enhanced nor correlated. We argue that this is due to a lack of shared expertise. Though in the case of music, Maess, Koelsch, Gunter, and Frederici (2001) found enhanced cortical activity in higher order areas also for novices. Notably, the authors used classical chords, which are familiar to Westerners.

Hasson and Malach (2006) suggested that ISC allows disentangling the cortex into two systems: areas where subjects process stereotypical responses to the external world and areas that may be linked to individual variation. Similarly, we propose that signs of processing and indices of understanding need to be distinguished. For instance, it is likely that music and/or action are at least partly processed in BA44, but in idiosyncratic ways. In order to link the enhanced cortical activity to shared understanding (as in mirror-neuron theories), one would also expect significant correlation. For instance, Lerner et al. (2011) found significant ISC in the primary auditory cortex on the level of words, whereas the correlation in higher auditory processing areas was sensitive to the length of the intact structure of spoken text. Thus, the more that could be understood (narrative) the higher auditory cortices were correlated between subjects. Hence, up to a certain level, novices process the stimuli in a similar manner; but irrespective of coherent multisensory integration processes, subsequent higher level processes can be idiosyncratic.

4.4 No correlation in the action observation network

During passive movement observation, a number of studies found convergent activity in fronto-parietal as well as occipito-temporal areas (see Grosbras et al., 2011). We did not find fronto-parietal areas to be correlating either across or within subjects (i.e. across conditions), as could be expected

from previous ISC studies (e.g. Hasson et al., [2004](#)) or from mirror-neuron studies measuring corresponding brain activity during action observation and action execution as the basis for action understanding (Rizzolatti & Sinigaglia, [2010](#)). We found, however, that novice subjects' BOLD responses were synchronised in the occipito-temporal areas. Presumably, somatosensory and emotional responses play an important role in watching dance performances (see also Arrighi et al., [2009](#)). Dance has become increasingly popular in cognitive science (for a review, see Bläsing et al., [2012](#)) and neuroimaging studies (for a review, see Sevdalis & Keller, [2011](#)), predominantly showing enhanced activity in fronto-parietal areas for dance experts who possess physical experience of the movements observed (Calvo-Merino, Glaser, Grèzes, Passingham, & Haggard, [2005](#); Calvo-Merino, Grèzes, Glaser, Passingham, & Haggard, [2006](#); Calvo-Merino, Jola, Glaser, & Haggard, [2008](#); Cross, Hamilton, & Grafton, [2006](#); Orgs, Dombrowski, Heil, & Jansen-Osmann, [2008](#); Pilgramm et al., [2010](#)). In an earlier transcranial magnetic stimulation study we showed that Bharatanatyam spectators require at least visual experience to enhance muscle-specific sensorimotor excitement (Jola, Abedian-Amiri, Kuppuswamy, Pollick, & Grosbras, [2012](#)). As stressed earlier, the emotional expressions have been found to be of a universal nature, but the dance and music in which the expressions are embedded are highly complex and unfamiliar. It is thus less surprising that the activity in areas associated with motor simulation and emotion recognition was uncorrelated and may be explained by a lack of shared motor or visual expertise between our novices.

It is possible that for synchronisation in the fronto-parietal network, expertise is required. However, Petrini et al. ([2011](#)) found reduced activity in areas of AV integration and action–sound representation in expert drummers when compared with novices, including fronto-temporal-parietal regions. Interestingly, while Cross and colleagues (Cross, Hamilton, Kraemer, Kelley, & Grafton, [2009a](#); Cross, Kraemer, Hamilton, Kelley, & Grafton, [2009b](#)) used music to accompany the movements that dancers learned, none of the studies on dance observation investigated the effect music has on the perception of movement. This is surprising, knowing that the mirror-neuron network is multimodal (e.g. Gazzola, Aziz-Zadeh, & Keysers, [2006](#); Kohler et al., [2002](#); Lahav, Saltzman, & Schlaug, [2007](#)) and dance is a complex, multidimensional stimulus consisting of a fluid mixture of body movement and sound. As the responses to naturally co-varying sound and actions have been found to be modified by expertise, future work is required comparing responses of novices and dance experts.

4.5 Merits of ISC

ISC indicates voxels that show significant correlation between subjects independent of the level of BOLD activity. ISC is therefore a potential complementary method along with other audio, visual, and AV integration designs (e.g. Beauchamp, [2005](#); Goebel & van Atteveldt, [2009](#); Kreifelts et al., [2010](#); Love et al., [2011](#)). Some known issues from conventional methods however remain while others are resolved. For instance, current scanners do not allow capturing individual neuronal activity within a voxel. Thus, neither ISC nor GLM conjunction analysis can distinguish between voxels where unisensory visual and auditory processes coexist and those where AV integration processes take place (Calvert and Thesen, [2004](#); Szameitat et al., [2011](#)). High-resolution scanning would allow identifying correlated activity of a smaller number of neurons across subjects, but it may reduce ISC as it also increases the effects of anatomical variability. It is thus important to investigate in designs that have greater statistical power but which are still applicable to exploratory approaches such as wavelet correlation (Lessa et al., [2011](#)).

Another issue relates to differences in attention: bimodal AV stimuli are generally coupled with an increased perceptual load in contrast to the unimodal stimulation of A or V (Kreifelts et al., [2010](#)). This may affect the level of BOLD response in sensory areas and may modify associated attentional resources. We argued that using ISC can partly circumvent this issue as ISC compares the changes in the BOLD responses over time, independent of the general level of activity. Nevertheless, in order to control for attention effects, such as a decline in attention over time, we randomised the presentation order of the three different sensory conditions (A, V, and AV). A separate analysis did not show any order effects. Furthermore, free viewing of a continuous stream of sensory information as in ISC is a more natural form of stimulation and can be considered to contain a higher social importance and be more entertaining for subjects throughout (see Jola & Grosbras, [2013](#)). Moreover, novelty has been considered an attention-influencing stimulus property in theories of aesthetic preferences (e.g. Berlyne, [1974](#)). With a narrative that is undecipherable for novices based on recognisable gestures, our stimuli are situated in a conflicting and thus arousing level of novelty.

Furthermore, the loud scanner noise presents a potential confound of conventional fMRI studies, but it is unlikely that it has significantly modified our results. First and foremost, the scanner noise was most notable in the visual only condition, where we found no auditory areas to be correlated between subjects. Thus, the scanner noise alone was not sufficient to drive ISC activation. Furthermore, we found no significant correlation in the insula, where activity is reportedly related to scanner noise (Schmidt et al., 2008). Moreover, MacSweeney et al. (2000) showed evidence that the STG activity is independent of scanner noise. In addition, effects of scanner noise on BOLD responses were found to be variable and thus uncorrelated between subjects (Ulmer et al., 1988).

Finally, highly controlled parameterised stimuli may have better allowed contrasting basic physical stimulus properties than the ISC with ecological valid stimuli. Ecologically valid stimuli indeed consist of a complex mixture of basic feature properties, but are however less arbitrary and closer to real life (less inference steps to be made), less dependent on pre-assumptions (e.g. such as on Haemodynamic Response Function), have fewer additional stimuli effects (e.g. created by artificial confounds), and prevent acquiring task strategies. We thus argue that natural viewing of complex stimuli of long duration as used here is well suited to studying human perception; that they provide an essential complement to artificial stimuli and laboratory tasks, and that our findings are a more genuine reflection of the implicit uni- and multisensory processes in the context of real life. A number of other studies on social interaction (e.g. Risko, Laidlaw, Freeth, Foulsham, & Kingstone, 2012) and AV integration (e.g. de Gelder & Bertelson, 2003; Kreifelts et al., 2010) have recently acknowledged the importance of ecological validity and the potential modification highly controlled but artificial stimuli can have on perceptual and cognitive processes. The future challenge is to build a model based on the combination of the two seemingly opposing approaches—the bottom-up approach where models of AV integration are built on results from artificial stimuli with simple cue combinations and the more top-down approach where AV integration is explored based on naturalistic stimuli.

5 Summary

ISC allows the exploration of sensory areas involved in natural viewing of long stimulus segments, i.e. >6 min. We found a correlation between subjects' voxel-based time course in previously reported uni- and multisensory areas (occipito-temporal) in early and late stages of sensory processing but we found that subjects' brain responses were not synchronised in higher order areas relevant for cognition, action, and/or emotion. Thus, this study highlights that by presenting a dance form unfamiliar to subjects, correspondences between subjects can be constraint onto the level of sensory AV processing. We also did not find cross-modal synchronisation (A to V), despite our stimuli showing a narrative of culturally specific choreographed movements to music. We thus situate our unfamiliar unedited but choreographed dance stimuli between edited feature films and random recordings of everyday scenes: we found less correlated areas synchronised than studies that used classically edited movies but more than for non-edited unchoreographed recordings of everyday situations. Our data support the idea that spectators' visual and auditory processes can be directed to some extent by choreographed movement and music without changes in the visual scene. Furthermore, ISC can show additional findings to conventional GLM analyses and should thus be considered a complementary tool to standard contrast analysis when exploring multisensory integration processes.

Acknowledgments. This study was supported by the Arts and Humanities Research Council (AHRC). Phil McAleer was supported by a grant from the Ministry of Defence. The authors thank Prof. Uri Hasson, Prof. Rainer Goebel, and Jukka-Pekka Kauppi for their helpful analysis suggestions. Furthermore, the authors thank Frances Crabbe for support with the scanning, Jen Todman for the icon picture, and the dance performer Dr Anna Kuppuswamy.

References

- Aguirre, G. K., Zarahn, E., & D'Esposito, M. (1998). The variability of human, BOLD hemodynamic responses. *NeuroImage*, 8(4), 360–369. doi:10.1006/nimg.1998.0367 ◀
- Arrighi, R., Marini, F., & Burr, D. (2009). Meaningful auditory information enhances perception of visual biological motion. *Journal of Vision*, 9(4), 25 (1–7). doi:10.1167/9.4.25 ◀
- Bartels, Z., & Zeki, S. (2004). The chronoarchitecture of the human brain—natural viewing conditions reveal a time-based anatomy of the brain. *NeuroImage*, 22, 419–433. doi:10.1016/j.neuroimage.2004.08.044 ◀

- Beauchamp, M. S. (2005). Statistical criteria in fMRI studies of multisensory integration. *Neuroinformatics*, 3(2), 93–113. doi:10.1385/NI:3:2:093 ◀
- Beauchamp, M. S., Argall, B. D., Bodurka, J., Duyn, J. H., & Martin, A. (2004). Unraveling multisensory integration: Patchy organization within human STS multisensory cortex. *Nature Neuroscience*, 7(11), 1190–1192. doi:10.1038/nn1333 ◀
- Benoit, M. M., Rajj, T., Lin, F. H., Jääskeläinen, I. P., & Stufflebeam, S. (2010). Primary and multisensory cortical activity is correlated with audiovisual percepts. *Human Brain Mapping*, 31(4), 526–538. doi:10.1093/cercor/bhq170 ◀
- Berlyne, D. E. (1974). *Studies in the new experimental aesthetics*. New York: Wiley. doi:10.1037/11164-000 ◀
- Bidet-Caulet, A., Voisin, J., Bertrand, O., & Fonlupt, P. (2005). Listening to a walking human activates the temporal biological motion area. *NeuroImage*, 28(1), 132–139. doi:10.1007/s00221-011-2620-4 ◀
- Bläsing, B., Calvo-Merino, B., Cross, E. S., Jola, C., Honisch, J., & Stevens, C. J. (2012). Neurocognitive control in dance perception and performance. *Acta Psychologica*, 139(2), 300–308. doi:10.1016/j.actpsy ◀
- Calvert, G. A. (2001). Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cerebral Cortex*, 11(12), 1110–1123. doi:10.1093/cercor/11.12.1110 ◀
- Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C. R., McGuire, P. K., David, A. S. (1997). Activation of auditory cortex during silent lipreading. *Science*, 276(5312), 593–596. doi:10.1126/science.276.5312.593 ◀
- Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current Biology*, 10(11), 649–657. doi:10.1016/S0960-9822(00)00513-3 ◀
- Calvert, G. A., & Thesen, T. (2004). Multisensory integration: Methodological approaches and emerging principles in the human brain. *Journal of Physiology Paris*, 98(1–3), 191–205. doi:10.1027/1618-3169.55.2.121 ◀
- Calvo-Merino, B., Glaser, D. E., Grèzes, J., Passingham, R. E., & Haggard, P. (2005). Action observation and acquired motor skills: An fMRI study with expert dancers. *Cerebral Cortex*, 15(8), 1243–1249. doi:10.1093/cercor/bhi007 ◀
- Calvo-Merino, B., Grèzes, J., Glaser, D. E., Passingham, R. E., & Haggard, P. (2006). Seeing or doing? Influence of visual and motor familiarity in action observation. *Current Biology*, 16(19), 1905–1910. doi:10.1007/s00426-010-0280-9 ◀
- Calvo-Merino, B., Jola, C., Glaser, D. E., & Haggard, P. (2008). Towards a sensorimotor aesthetics of performing art. *Consciousness and Cognition*, 17(3), 911–922. doi:10.3389/fnhum.2011.00102 ◀
- Cross, E. S., Hamilton, A. F., & Grafton, S. T. (2006). Building a motor simulation de novo: Observation of dance by dancers. *NeuroImage*, 31(3), 1257–1267. doi:10.1016/j.neuroimage.2006.01.033 ◀
- Cross, E. S., Hamilton, A. F., Kraemer, D. J., Kelley, W. M., & Grafton, S. T. (2009a). Dissociable substrates for body motion and physical experience in the human action observation network. *European Journal of Neuroscience*, 30(7), 1383–1392. doi:10.1111/j.1460-9568.2009.06941.x ◀
- Cross, E. S., Kraemer, D. J., Hamilton, A. F., Kelley, W. M., & Grafton, S. T. (2009b). Sensitivity of the action observation network to physical and observational learning. *Cerebral Cortex*, 19(2), 315–326. doi:10.1093/cercor/bhn083 ◀
- de Gelder, B., & Bertelson, P. (2003). Multisensory integration, perception and ecological validity. *TRENDS in Cognitive Sciences*, 7(10), 460–467. doi:10.1016/j.tics.2003.08.014 ◀
- de Gelder, B., Snyder, J., Greve, D., Gerard, G., & Hadjikhani, N. (2004). Fear fosters flight: A mechanism for fear contagion when perceiving emotion expressed by a whole body. *PNAS*, 101(47), 16701–16706. doi:10.1073/pnas.0407042101 ◀
- Downing, P. E., & Peelen, M. V. (2011). The role of occipitotemporal body-selective regions in person perception. *Cognitive Neuroscience*, 2(3–4), 186–203. doi:10.1002/mds.23736 ◀
- Ethofer, T., Pourtois, G., & Wildgruber, D. (2006). Investigating audiovisual integration of emotional signals in the human brain. *Progress in Brain Research*, 156, 345–361. doi:10.1162/jocn.2009.21099 ◀
- Fenlon, J., Denmark, T., Campbell, R., & Woll, B. (2007). Seeing sentence boundaries. *Sign Language and Linguistics*, 10(2), 177–200. doi:10.1075/sll.10.2.06fen ◀
- Forman, S. D., Cohen, J. D., Fitzgerald, M., Eddy, W. F., Mintun, M. A., & Noll, D. C. (1995). Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): Use of a cluster-size threshold. *Magnetic Resonance in Medicine*, 33(5), 636–647. doi:10.1002/mrm.1910330508 ◀
- Friston, K. J. (2005). Models of brain function in neuroimaging. *Annual Review of Psychology*, 56, 57–87. doi:10.1146/annurev.psych.56.091103.070311 ◀
- Friston, K. J., Penny, W., & Glaser, D. E. (2005). Conjunction revisited. *NeuroImage*, 25(3), 661–667. doi:10.1002/hbm.20242 ◀
- Furman, O., Dorfman, N., Hasson, U., Davachi, L., & Dudai, Y. (2007). They saw a movie: Long-term memory for an extended audiovisual narrative. *Learning & Memory*, 14(6), 457–467. doi:10.1101/lm.550407 ◀

- Gauthier, I., Tarr, M. J., Moylan, J., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). The fusiform “face area” is part of a network that processes faces at the individual level. *Journal of Cognitive Neuroscience*, *12*(3), 495–504. [doi:10.1037/0096-1523.28.2.431](https://doi.org/10.1037/0096-1523.28.2.431) ◀
- Gazzola, V., Aziz-Zadeh, L., & Keysers, Chr. (2006). Empathy and the somatotopic auditory mirror system in humans. *Current Biology*, *16*(18), 1824–1829. [doi:10.1016/j.cub.2006.07.072](https://doi.org/10.1016/j.cub.2006.07.072) ◀
- Goebel, R., Esposito, F., & Formisano, E. (2006). Analysis of functional image analysis contest (FIAC) data with brainvoyager QX: From single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. *Human Brain Mapping*, *27*(5), 392–401. [doi:10.1002/hbm.20249](https://doi.org/10.1002/hbm.20249) ◀
- Goebel, R., & van Atteveldt, N. (2009). Multisensory functional magnetic resonance imaging: A future perspective. *Experimental Brain Research*, *198*(2–3), 153–164. [doi:10.1186/1471-2202-11-11](https://doi.org/10.1186/1471-2202-11-11) ◀
- Grahn, J. A., Henry, M. J., & McAuley, J. D. (2011). fMRI investigation of cross-modal interactions in rhythm perception: Audition primes vision, but not vice versa. *NeuroImage*, *54*(2), 1231–1243. [doi:10.1016/j.neuroimage.2010.09.033](https://doi.org/10.1016/j.neuroimage.2010.09.033) ◀
- Grafton, S. T. (2009). Embodied cognition and the simulation of action to understand others. *Annals of the New York Academy of Sciences*, *1156*, 97–117. [doi:10.1111/j.1749-6632.2009.04425.x](https://doi.org/10.1111/j.1749-6632.2009.04425.x) ◀
- Grill-Spector, K., & Malach, R. (2004). The human visual cortex. *Annual Review of Neuroscience*, *27*, 649–677. [doi:10.1146/annurev.neuro.27.070203.144220](https://doi.org/10.1146/annurev.neuro.27.070203.144220). ◀
- Grosbras, M. H., Beaton, S., & Eickhoff, S. B. (2011). Brain regions involved in human movement perception: A quantitative voxel-based meta-analysis. *Human Brain Mapping*, *33*(2), 431–454. [doi:10.1002/hbm.21222](https://doi.org/10.1002/hbm.21222) ◀
- Handwerker, D. A., Ollinger, J. M., & D’Esposito, M. (2004). Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *NeuroImage*, *21*(4), 1639–1651. [doi:10.1016/j.neuroimage.2009.11.014](https://doi.org/10.1016/j.neuroimage.2009.11.014) ◀
- Hasson, U., Avidan, G., Gelbard, H., Vallines, I., Harel, M., Minshew, N., & Behrmann, M. (2009a). Shared and idiosyncratic cortical activation patterns in autism revealed under continuous real-life viewing conditions. *Autism Research*, *2*(4), 220–231. [doi:10.1016/j.visres.2012.11.002](https://doi.org/10.1016/j.visres.2012.11.002) ◀
- Hasson, U., Furman, O., Clark, D., Dudai, Y., & Davachi, L. (2008a). Enhanced intersubject correlations during movie viewing correlate with successful episodic encoding. *Neuron*, *57*(3), 452–462. [doi:10.3389/fnhum.2012.00248](https://doi.org/10.3389/fnhum.2012.00248) ◀
- Hasson, U., Landesman, O., Knappmeyer, B., Vallines, I., Rubin, N., & Heeger, D. (2008b). Neurocinematics: The neuroscience of films. *Projections: The Journal for Movies and Mind*, *2*(1), 1–26. [doi:10.3167/proj.2008.020102](https://doi.org/10.3167/proj.2008.020102) ◀
- Hasson, U., & Malach, R. (2006). Human brain activation during viewing of dynamic natural scenes. *Novartis Foundation Symposium*, *270*, 203–212. [doi:10.1155/2012/375148](https://doi.org/10.1155/2012/375148) ◀
- Hasson, U., Malach, R., & Heeger, D. J. (2009b). Reliability of cortical activity during natural stimulation. *Trends in Cognitive Sciences*, *14*(1), 40–48. [doi:10.1371/journal.pbio.1001462](https://doi.org/10.1371/journal.pbio.1001462) ◀
- Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science*, *303*(5664), 1634–1640. [doi:10.1126/science.1089506](https://doi.org/10.1126/science.1089506) ◀
- Hasson, U., Yang, E., Vallines, I., Heeger, D. J., & Rubin, N. (2008c). A hierarchy of temporal receptive windows in human cortex. *Journal of Neuroscience*, *28*(10), 2539–2550. [doi:10.1523/JNEUROSCI.3684-10](https://doi.org/10.1523/JNEUROSCI.3684-10) ◀
- Hein, G., & Knight, R. T. (2008). Superior temporal sulcus—It’s my area: Or is it? *Cognitive Neuroscience*, *20*(12), 2125–2136. [doi:10.3410/f.13254956.14607054](https://doi.org/10.3410/f.13254956.14607054) ◀
- Hejnar, M. P., Kiehl, K. A., & Calhoun, V. D. (2007). Interparticipant correlations: A model free fMRI analysis technique. *Human Brain Mapping*, *28*(9), 860–867. [doi:10.1002/hbm.20321](https://doi.org/10.1002/hbm.20321) ◀
- Hejmadi, A., Davidson, R. J., & Paul, R. (2000). Exploring Hindu Indian emotion expressions evidence for accurate recognition by Americans and Indians. *Psychological Science*, *11*(4), 183–187. [doi:10.1093/acprof:oso/9780195373585.003.0029](https://doi.org/10.1093/acprof:oso/9780195373585.003.0029) ◀
- Jääskeläinen, I. P., Koskentalo, K., Balk, M. H., Autti, T., Kauramäki, J., Pomren, C., & Sams, M. (2008). Inter-subject synchronization of prefrontal cortex hemodynamic activity during natural viewing. *Open Neuroimage Journal*, *2*, 14–19. [doi:10.3389/fnhum.2012.00298](https://doi.org/10.3389/fnhum.2012.00298) ◀
- Jola, C., Abedian-Amiri, A., Kuppuswamy, A., Pollick, F. E., & Grosbras, M. H. (2012). Motor simulation without motor expertise: Enhanced corticospinal excitability in visually experienced dance spectators. *PLoS One*, *7*(3), e33343. [doi:10.1371/journal.pone.0033343](https://doi.org/10.1371/journal.pone.0033343) ◀
- Jola, C., Davis, A., & Haggard, P. (2011). Proprioceptive integration and body representation: Insights into dancers’ expertise. *Experimental Brain Research*, *213*(2–3), 257–265. [doi:10.1016/j.actpsy.2011.12.005](https://doi.org/10.1016/j.actpsy.2011.12.005) ◀
- Jola, C., & Grosbras, M.-H. (2013). In the here and now: Enhanced motor corticospinal excitability in novices when watching live compared to video recorded dance. *Cognitive Neuroscience*. [doi:10.1080/17588928.2013.776035](https://doi.org/10.1080/17588928.2013.776035) ◀

- Kauppi, J. P., Jääskeläinen, I. P., Sams, M., & Tohka, J. (2010). Inter-subject correlation of brain hemodynamic responses during watching a movie: Localization in space and frequency. *Frontiers in Neuroinformatics*, 4(5). doi:10.3389/fninf.2010.00005 ◀
- Kilian-Hütten, N., Vroomen, J., & Formisano, E. (2011). Brain activation during audiovisual exposure anticipates future perception of ambiguous speech. *NeuroImage*, 57(4), 1601–1607. doi:10.1016/j.neuroimage.2011.05.043 ◀
- Klemen, J., & Chambers, C. D. (2011). Current perspectives and methods in studying neural mechanisms of multisensory interactions. *Neuroscience and Biobehavioral Reviews*, 36(1), 111–133. doi:10.1155/2012/720278 ◀
- Koelsch, S. (2011). Toward a neural basis of music perception—a review and updated model. *Frontiers in Psychology*, 2, 110. doi:10.3389/fpsyg.2011.00110 ◀
- Kohler, E., Keysers, C., Umiltà, M. A., Fogassi, L., Gallese, V., & Rizzolatti, G. (2002). Hearing sounds, understanding actions: Action representation in mirror neurons. *Science*, 297(5582), 846–848. doi:10.1073/pnas.1205553109 ◀
- Kreifelts, B., Ethofer, T., Grodd, W., Erb, M., & Wildgruber, D. (2007). Audiovisual integration of emotional signals in voice and face: An event-related fMRI study. *NeuroImage*, 37(4), 1445–1456. doi:10.1016/j.neuroimage.2007.06.020 ◀
- Kreifelts, B., Ethofer, T., Huberle, E., Grodd, W., & Wildgruber, D. (2010). Association of trait emotional intelligence and individual fMRI-activation patterns during the perception of social signals from voice and face. *Human Brain Mapping*, 31(7), 979–991. doi:10.1002/hbm.20913 ◀
- Krumhansl, C. L., & Schenck, D. L. (1997). Can dance reflect the structural and expressive qualities of music? A perceptual experiment on Balanchine's choreography of Mozart's Divertimento no. 15. *Musicae Scientiae*, 1(1), 63–85. doi:10.1111/j.1469-8986.1969 ◀
- Laurienti, P. J., Burdette, J. H., Wallace, M. T., Yen, Y., Field, A. S., & Stein, B. E. (2002). Deactivation of sensory-specific cortex by polymodal stimuli. *Journal of Cognitive Neuroscience*, 14(3), 420–429. doi:10.1523/JNEUROSCI.0910-09.2009 ◀
- Lahav, A., Saltzman, E., & Schlaug, G. (2007). Action representation of sound: Audiomotor recognition network while listening to newly acquired actions. *Journal of Neuroscience*, 27(2), 308–314. doi:10.1523/JNEUROSCI.4822-06.2007 ◀
- Lancaster, J. L., Woldorff, M. G., Parsons, L. M., Liotti, M., Freitas, C. S., Rainey, L., Fox, P. T. (2000). Automated Talairach atlas labels for functional brain mapping. *Human Brain Mapping*, 10(3), 120–131. ◀
- Le, T. H., Pardo, J. V., & Hu, X. (1998). 4 T-fMRI study of nonspatial shifting of selective attention: Cerebellar and parietal contributions. *Journal of Neurophysiology*, 79(3), 1535–1548. doi:10.1016/j.neuroimage.2005.10.032 ◀
- Lerner, Y., Honey, C. J., Silbert, L. J., & Hasson, U. (2011). Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. *Journal of Neuroscience*, 31(8), 2906–2915. doi:10.1523/JNEUROSCI.3684 ◀
- Lessa, P. S., Sato, J. R., Cardoso, E. F., Neto, C. G., Valadares, A. P., & Amaro, E. (2011). Wavelet correlation between subjects: A time-scale data driven analysis for brain mapping using fMRI. *Journal of Neuroscience Methods*, 194(2), 350–357. doi:10.1016/j.jneumeth.2010.09.005 ◀
- Li, Y., Wang, G., Long, J., Yu, Z., Lian, G., Li, Z., & Sun, P. (2011). Reproducibility and discriminability of brain patterns of semantic categories enhanced by congruent audiovisual stimuli. *PLoS One*, 6(6), e20801. doi:10.1371/journal.pone.0020708 ◀
- Love, S. A., Pollick, F. E., & Latinus, M. (2011). Cerebral correlates and statistical criteria of cross-modal face and voice integration. *Seeing and Perceiving*, 24(4), 351–367. doi:10.1371/journal.pone.0019165 ◀
- Love, S. A., Pollick, F. E., & Petrini, K. (2012). Effects of experience, training and expertise on multisensory perception: Investigating the link between brain and behaviour. *Lecture Notes in Computer Science: Proceedings on Cognitive Behavioural Systems*, 7403, 304–320. doi:10.1007/978-3-642-34584-5_27 ◀
- MacSweeney, M., Amaro, E., Calvert, G. A., Campbell, R., David, A. S., McGuire, P., Brammer, M. J. (2000). Silent speechreading in the absence of scanner noise: An event-related fMRI study. *Neuroreport*, 11(8), 1729–1733. ◀
- MacSweeney, M., Campbell, R., Woll, B., Giampietro, V., David, A. S., McGuire, P. K., Brammer, M. J. (2004). Dissociating linguistic and nonlinguistic gestural communication in the brain. *NeuroImage*, 22(4), 1605–1618. doi:10.1016/j.neuroimage.2004.03.015 ◀
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264(5588), 746–748. doi:10.1073/pnas.0804275105 ◀
- McNamara, A., Buccino, G., Menz, M. M., Gläscher, J., Wolbers, T., Baumgärtner, A., & Binkofski, F. (2008). Neural dynamics of learning sound–action associations. *PLoS One*, 3(12), e3845. doi:10.1016/j.physbeh.2008.06.011 ◀
- Maess, B., Koelsch, S., Gunter, T. C., & Frederici, A. D. (2001). Musical syntax is processed in Broca's area: An MEG study. *Nature Neuroscience*, 4(5), 540–545. doi:10.3758/CABN.8.3.318 ◀

- Meyer, G. F., Greenlee, M., & Wuerger, S. (2011). Interactions between auditory and visual semantic stimulus classes: Evidence for common processing networks for speech and body actions. *Journal of Cognitive Neuroscience*, 23(9), 2291–2308. doi:10.1162/jocn.2010.21575 ◀
- Möttönen, R., Calvert, G. A., Jääskeläinen, I. P., Matthews, P. M., Thesen, T., Tuomainen, J., & Sams, M. (2006). Perceiving identical sounds as speech or non-speech modulates activity in the left posterior superior temporal sulcus. *NeuroImage*, 30(2), 563–593. doi:10.1016/j.neuroimage.2005.10.002 ◀
- Nath, A. R., & Beauchamp, M. S. (2011). Dynamic changes in superior temporal sulcus connectivity during perception of noisy audiovisual speech. *Journal of Neuroscience*, 31(5), 1704–1714. doi:10.1523/JNEUROSCI.2605-11.2011 ◀
- Navarra, J., & Soto-Faraco, S. (2005). Hearing lips in a second language: Visual articulatory information enables the perception of second language sounds. *Psychological Research*, 71(1), 4–12. doi:10.1016/j.actpsy.2008.08.004 ◀
- Orgs, G., Dombrowski, J. H., Heil, M., & Jansen-Osmann, P. (2008). Expertise in dance modulates alpha/beta event-related desynchronization during action observation. *European Journal of Neuroscience*, 27(12), 3380–3384. doi:10.1111/j.1467-7687.2010.00991.x ◀
- Ou, W., Raji, T., Lin, F.-H., Golland, P., & Hämäläinen, M. (2009). Modeling adaptation effects in fMRI analysis. *Medical Image Computing and Computer-Assisted Intervention*, 12(Pt 1), 1009–1017. doi:10.1007/978-3-642-04268-3_124 ◀
- Petrini, K., Pollick, F. E., Dahl, S., McAleer, P., McKay, L. S., Rocchesso, D., Puce, A. (2011). Action expertise reduces brain activity for audiovisual matching actions: An fMRI study with expert drummers. *NeuroImage*, 56(3), 1480–1492. doi:10.1016/j.neuroimage.2011.03.009 ◀
- Pilgramm, S., Lorey, B., Stark, R., Munzert, J., Vaitl, D., & Zentgraf, K. (2010). Differential activation of the lateral premotor cortex during action observation. *BMC Neuroscience*, 11, 89. doi:10.1186/1471-2202-11-89 ◀
- Pillai, S. (2002). Rethinking global Indian dance through local eyes: The contemporary Bharatanatyam scene in Chennai. *Dance Research Journal*, 34(2), 14–29. doi:10.1080/14647890903568305 ◀
- Raji, T., Uutela, K., & Hari, R. (2000). Audiovisual integration of letters in the human brain. *Neuron*, 28(2), 617–625. doi:10.1016/S0896-6273(00)00138-0 ◀
- Reale, R. A., Calvert, G. A., Thesen, T., Jenison, R. L., Kawasaki, H., Oya, H., Brugge, J. F. (2007). Auditory–visual processing represented in the human superior temporal gyrus. *Neuroscience*, 145(1), 162–84. doi:10.1016/j.bandl.2008.10.005 ◀
- Reason, M., & Reynolds, D. (2010). Kinesthesia, empathy, and related pleasures: An inquiry into audience experiences of watching dance. *Dance Research Journal*, 42(2), 49–75. doi:10.3366/drs.2011.0019 ◀
- Risko, E. F., Laidlaw, K., Freeth, M., Foulsham, T., & Kingstone, A. (2012). Social attention with real versus reel stimuli: Toward an empirical approach to concerns about ecological validity. *Frontiers in Human Neuroscience*, 6, 143. doi:10.3389/fnhum.2012.00143 ◀
- Rizzolatti, G., & Sinigaglia, C. (2010). The functional role of the parieto-frontal mirror circuit: Interpretations and misinterpretations. *Nature Review Neuroscience*, 11(4), 264–274. doi:10.1038/nrn2805 ◀
- Schmidt, C. F., Zaehle, T., Meyer, M., Geiser, E., Boesiger, P., & Jancke, L. (2008). Silent and continuous fMRI scanning differentially modulate activation in an auditory language comprehension task. *Human Brain Mapping*, 29(1), 46–56. doi:10.1371/journal.pone.0054273 ◀
- Sekiyama, K., Kanno, I., Miura, S., & Sugita, Y. (2003). Auditory–visual speech perception examined by fMRI and PET. *Neuroscience Research*, 47(3), 277–287. doi:10.1016/S0168-0102(03)00214-1 ◀
- Sevdalis, V., & Keller, P. E. (2011). Captured by motion: Dance, action understanding, and social cognition. *Brain and Cognition*, 77(2), 231–236. doi:10.3389/fnhum.2011.00102 ◀
- Szameitat, A. J., Schubert, T., & Müller, H. J. (2011). How to test for dual-task-specific effects in brain imaging studies—an evaluation of potential analysis methods. *NeuroImage*, 54(3), 1765–1773. doi:10.1037a0025816 ◀
- Szyck, G. R., Tausche, P., & Münte, T. F. (2008). A novel approach to study audiovisual integration in speech perception: Localizer fMRI and sparse sampling. *Brain Research*, 1220, 142–149. doi:10.1016/j.brainres.2007.08.027 ◀
- Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain: 3-dimensional proportional system—an approach to cerebral imaging*. New York: Thieme Medical Publishers. doi.org/10.1016/S1361-8415(01)00043-3 ◀
- Ulmer, J. L., Biswal, B. B., Yetkin, F. Z., Mark, L. P., Mathews, V. P., Prost, R. W., Daniels, D. L. (1998). Cortical activation response to acoustic echo planar scanner noise. *Journal of Computer Assisted Tomography*, 22(1), 111–119. ◀
- Vatsyayan, K. (1963). Notes on the relationship of music and dance in India. *Ethnomusicology*, 7(1), 33–38. doi:10.2307/924145 ◀
- Vines, B. W., Krumhansl, C. L., Wanderley, M. M., Dalca, I. M., & Levitin, D. J. (2011). Music to my eyes: Cross-modal interactions in the perception of emotions in musical performance. *Cognition*, 118(2), 157–170. doi:10.1016/j.cognition.2010.11.010 ◀

-
- Vines, B. W., Krumhansl, C. L., Wanderley, M. M., & Levitin, D. J. (2006). Cross-modal interactions in the perception of musical performance. *Cognition*, *101*(1), 80–113. [doi:10.1016/j.cognition.2005.09.003](https://doi.org/10.1016/j.cognition.2005.09.003) ◀
- Wolf, I., Dziobek, I., & Heekeren, H. R. (2010). Neural correlates of social cognition in naturalistic settings: A model-free analysis approach. *NeuroImage*, *49*(1), 894–904. [doi:10.1016/j.neuroimage.2009.08.060](https://doi.org/10.1016/j.neuroimage.2009.08.060) ◀
- Wright, T. M., Pelphrey, K. A., Allison, T., McKeown, M. J., & McCarthy, G. (2003). Polysensory interactions along lateral temporal regions evoked by audiovisual speech. *Cerebral Cortex*, *13*(10), 1034–1043. [doi:10.3758/s13414-012-0375-z](https://doi.org/10.3758/s13414-012-0375-z) ◀
- Zacks, J. M., Braver, T. S., Sheridan, M. A., Donaldson, D. I., Snyder, A. Z., Ollinger, J. M., Raichle, M. E. (2001). Human brain activity time-locked to perceptual event boundaries. *Nature Neuroscience*, *4*(6), 651–655. [doi:10.1002/wcs.133](https://doi.org/10.1002/wcs.133) ◀
- Zacks, J. M., Speer, N. K., Swallow, K. M., & Maley, C. J. (2010). The brain's cutting-room floor: Segmentation of narrative cinema. *Frontiers in Human Neuroscience*, *4*, 168. [doi:10.1007/s11098-010-9562-8](https://doi.org/10.1007/s11098-010-9562-8) ◀