

How music alters a kiss: superior temporal gyrus controls fusiform–amygdalar effective connectivity

Corinna Pehrs,^{1,2,3,4} Lorenz Deserno,^{5,6} Jan-Hendrik Bakels,^{1,7} Lorna H. Schlochtermeyer,^{1,2,3,8} Hermann Kappelhoff,^{1,7} Arthur M. Jacobs,^{1,2,3} Thomas Hans Fritz,^{6,9,10} Stefan Koelsch,^{1,3,6} and Lars Kuchinke^{1,2,3,11}

¹Cluster of Excellence ‘Languages of Emotion’, Freie Universität Berlin, Habelschwerdter Allee 45, 14195 Berlin, Germany, ²Department of Education and Psychology, Freie Universität Berlin, Habelschwerdter Allee 45, 14195 Berlin, Germany, ³Dahlem Institute for Neuroimaging of Emotion, Freie Universität Berlin, Habelschwerdter Allee 45, 14195 Berlin, Germany, ⁴Department of Psychology, Stanford University, Stanford, CA 94305, USA, ⁵Department of Psychiatry and Psychotherapy, Campus Charité Mitte, Charité-Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany, ⁶Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstrasse 1a, 04103 Leipzig, Germany, ⁷Department of Philosophy and Humanities, Freie Universität Berlin, Habelschwerdter Allee 45, 14195 Berlin, Germany, ⁸Department of Psychology, Princeton University, Princeton, NJ 08540, USA, ⁹Department of Nuclear Medicine, University of Leipzig, Liebigstrasse 18, 04103 Leipzig, Germany, ¹⁰Institute for Psychoacoustics and Electronic Music (IPEM), Blandijnberg 2, B-9000 Ghent, Belgium and ¹¹Department of Psychology, Experimental Psychology and Methods, Ruhr-Universität Bochum, Universitätsstraße 150, 44801 Bochum, Germany

While watching movies, the brain integrates the visual information and the musical soundtrack into a coherent percept. Multisensory integration can lead to emotion elicitation on which soundtrack valences may have a modulatory impact. Here, dynamic kissing scenes from romantic comedies were presented to 22 participants (13 females) during functional magnetic resonance imaging scanning. The kissing scenes were either accompanied by happy music, sad music or no music. Evidence from cross-modal studies motivated a predefined three-region network for multisensory integration of emotion, consisting of fusiform gyrus (FG), amygdala (AMY) and anterior superior temporal gyrus (aSTG). The interactions in this network were investigated using dynamic causal models of effective connectivity. This revealed bilinear modulations by happy and sad music with suppression effects on the connectivity from FG and AMY to aSTG. Non-linear dynamic causal modeling showed a suppressive gating effect of aSTG on fusiform–amygdalar connectivity. In conclusion, fusiform to amygdala coupling strength is modulated via feedback through aSTG as region for multisensory integration of emotional material. This mechanism was emotion-specific and more pronounced for sad music. Therefore, soundtrack valences may modulate emotion elicitation in movies by differentially changing preprocessed visual information to the amygdala.

Keywords: fMRI; non-linear DCM; multisensory integration; superior temporal gyrus; emotion

INTRODUCTION

While watching a movie, spectators make use of the underlying film music as a source of emotional information (Cohen, 2001; Kuchinke *et al.*, 2013). The evaluation of the same visual kissing scene can be altered by emotionally loaded music. A kiss is still a kiss, but a happy music score may indicate a positive development of the scene, whereas a sad or melancholy music score may refer to a melodramatic ending. By integrating visual and auditory information, film parallels important aspects of everyday emotional experiences. Both music (e.g. Koelsch, 2010) and film in particular (Gross and Levenson, 1995) are known for their advantageous capacity to elicit strong emotions. Hence, combining film and music as experimental stimuli addresses the mechanisms involved in multisensory integration and emotion processing.

Multisensory integration and emotion processing

The brain is highly adapted to integrate sensory inputs from different perceptual channels. Combining auditory and visual information alters the perception of unimodal presentation. Experiments conducted by

de Gelder and Vroomen (2000) demonstrated that individuals show higher accuracy and faster reaction times in classifying emotional facial expressions when accompanied by congruent emotional prosody (see also Pourtois *et al.*, 2000; Ethofer *et al.*, 2006). However, incongruent audiovisual presentation may lead to perceptual illusions when, for example, the lip movements do not match the spoken syllable (cf. the McGurk effect; McGurk and McDonald, 1976). The biased perception of emotions in sound and facial expressions due to the emotion presented in the other modality is thus labeled the ‘emotional McGurk effect’ (de Gelder and Bertelson, 2003).

Neural correlates of multisensory emotion processing

Neuroimaging research has identified some candidate regions for a network of multisensory integration of emotional content. Dolan *et al.* (2001) reported increased hemodynamic responses in fusiform gyrus (FG) and amygdala (AMY) when fearful faces were presented together with a congruent vocal expression of emotion. This additive effect in FG and AMY was replicated several times (Baumgartner *et al.*, 2006; Ethofer *et al.*, 2006; Eldar *et al.*, 2007). Dolan *et al.* (2001) assumed greater connectivity between AMY and FG during congruent conditions than during incongruent conditions as a possible explanation.

Embedded in the framework of heteromodal convergence zones (Damasio, 1989; Mesulam, 1998), the superior temporal gyrus (STG) has been identified as a region for multisensory integration (e.g. Calvert and Thesen, 2004; Pourtois *et al.*, 2005; Kreifelts *et al.*, 2007). Of particular note is a likely emotional vs non-emotional differentiation along the STG. A study by Robins *et al.* (2009) revealed

Received 20 December 2012; Revised 17 September 2013; Accepted 31 October 2013

Advance Access publication 2 December 2013

The authors thank Tila-Tabeba Brink, Maria Tsaoussoglou and Hilda Hohl for help during data acquisition. They thank Stavros Skouras for help with data analysis and Andrea Samson and Valeria Manera for fruitful discussions. Furthermore, they thank Ursula Beerman and Dar Meshi for comments on a previous version of this manuscript. This work was funded by the German Research Foundation (DFG, Cluster of Excellence ‘Languages of Emotion’, EXC302). The study sponsor had no influence on study design, on the collection, analysis and interpretation of data, on the writing of the report and on the decision to submit the article for publication.

Correspondence should be addressed to Corinna Pehrs, Cluster Languages of Emotion, Freie Universität Berlin, Habelschwerdter Allee 45, 14195 Berlin, Germany. E-mail: corinna.pehrs@fu-berlin.de

that the integration of emotional audiovisual material is neuroanatomically dissociable from audiovisual integration *per se* and located in the anterior part of the STG (aSTG). Based on these findings, the STG was proposed as a heteromodal structure, which mediates modulatory influences (Calvert and Thesen, 2004).

To date, only a few studies have applied connectivity analyses on multisensory-emotional integration paradigms. For example, changes in connectivity between the posterior (pSTG) and unimodal sensory areas for visual and auditory input were observed during congruent conditions using psychophysiological interactions (Macaluso *et al.*, 2000; Kreifelts *et al.*, 2007). This finding was interpreted as a backward modulation from multisensory (pSTG) to unimodal sensory areas but might also be indicative of more complex modulatory influences. As heteromodal area, the aSTG is a candidate region, which gates connectivity between two other regions during multisensory integration of emotional material. Such a gating structure has been shown in a more recent dynamic causal modeling (DCM) cross-modal functional magnetic resonance imaging (fMRI) study on fusiform–amygdalar connectivity (Müller *et al.*, 2012) using static pictures of facial expressions and simple sounds (happy, fearful, sad, neutral).

This study

Gating effects conveyed by aSTG might underlie the observation that music is able to strongly impact the perception of complex dynamic visual stimuli (Vitouch, 2001; Boltz, 2005). This study was designed to examine how music changes the perception of popular emotional cinema movie sequences and, in particular, to test the hypothesis that aSTG mediates perceptual–emotional changes under naturalistic viewing conditions. Therefore, emotionally intense kissing scenes taken from romantic comedies were combined with happy and sad music. A highly standardized auditory stimulus set was developed containing carefully selected happy and sad music pieces matched for pleasantness/unpleasantness, instrumentation, style and tempo.

Effects of emotional music during movie perception were expected to activate a multisensory integration network including aSTG, FG and AMY. We hypothesized that the emotional effects of music would change the dynamics between the nodes of this network such that (i) aSTG receives information on emotional qualities from FG and AMY and (ii) aSTG controls fusiform–amygdalar connectivity.

MATERIALS AND METHODS

Participants

Twenty-two right-handed healthy subjects (13 female, 9 males) with a mean age of 26.68 (s.d. = 6.99, min = 20, max = 51) participated in the study. Participants did not have any neurological or psychiatric disorders. Written informed consent was obtained, and the study was approved by the ethics committee of the German Psychological Society. Participants either received course credit or 10€/h for participation (25€ in total).

Visual stimuli

Sixty-four kissing scenes were selected based on film analytic methods referring to an electronically based media analysis of expressive movement images (eMAEX; Kappelhoff and Bakels, 2011) from Hollywood movies of the romantic comedy genre (produced between 1995 and 2009; Supplementary Table S1). An independent sample of 25 individuals (16 females, 9 males, mean age = 30.2, s.d. = 10.06) rated all movies presented without sound on a scale from 1 to 7 for valence, arousal, prototypicality for kissing scenes, familiarity and liking of the actors. Based on these ratings, a set of the most homogenously rated 24 film clips was created. These film clips had an average length of 64.45 s (s.d. = 7.84). All actors or actresses appeared only once. Only scenes

were selected, in which the actors and actresses were liked (female actors $M = 4.78$, s.d. = 0.52; male actors $M = 4.51$, s.d. = 0.52). Furthermore, they were experienced as physiologically arousing with a mean of 3.95 (s.d. = 0.46) and as prototypical for kissing scenes in feature films ($M = 4.72$, s.d. = 0.55). Their familiarity ranged from 1.07 to 4 ($M = 2.36$, s.d. = 0.52). The film clips were processed and rendered using the film editing Software Adobe Premiere Pro CS5.5 (Adobe System Incorporated®, San Jose, CA, USA). Some film clips were scaled in order to reach a uniform screen 4:3 format and to avoid optical distortions. The film clips were faded in and out using a 0.8 s fade to and from a black screen. The fading properties of the music were matched according to the video shades. For presentation purposes, all film clips were saved with a resolution of 600 × 800 pixels.

Auditory stimuli

To assure orthogonality of emotional experimental conditions, happy and sad music was chosen (Supplementary Table S2). First, pairs of musical pieces of the opposite affective tone (happy–sad) were matched in instrumentation and musical genre. All musical stimuli were instrumental excerpts, all sad stimuli in minor and all happy stimuli in major. Because happy music is usually faster than sad music, leading to physiological effects such as increased heart rate and increased breathing rate due to the tempo of stimuli (Bernardi *et al.*, 2006), we controlled happy and sad stimuli for differences in tempo characteristics: for each happy–sad pair, an electronic beat was composed and added to the original music using Ableton Live 8 (Ableton Inc., Berlin) and a drum and percussion sample database (Ableton Inc.). Thus, both happy and sad pieces of a happy–sad pair were overlaid with an acoustically identical beat track (tempo variations that led to deviations of the original versions from the beat track were removed using the warping and time-stretching functions in Ableton Live 8). For rendering, the volume of the percussive rhythms was set to 3 dB below the volume of the musical excerpts during the merging process. Finally, the waveforms of all auditory stimuli were RMS-normalized (root mean square) using the digital audio software Adobe Audition CS5.5 (Adobe System Incorporated®). Behavioral data obtained during the fMRI scanning session showed the appropriateness of the stimuli to elicit happy and sad emotions with similar levels of arousal (Table 1).

Table 1 Behavioral ratings of the kissing scenes used for the fMRI experiment ($N = 22$) for the mean effect of music (happy and sad vs mute condition) and the music emotion effect (happy vs sad) with standard error (s.e.) in brackets

	Music	No music	Paired <i>t</i> -test	
	Mean (s.e.)	Mean (s.e.)	<i>t</i> -value	<i>P</i> -value
Valence	1.00 (0.12)	0.84 (0.13)	1.53	0.139
Arousal	3.51 (0.21)	3.23 (0.22)	2.06	0.052
	Happy	Sad	Paired <i>t</i> -test	
	Mean (s.e.)	Mean (s.e.)	<i>t</i> -value	<i>P</i> -value
Valence	1.13 (0.15)	0.87 (0.11)	2.04	0.054
Arousal	3.58 (0.22)	3.44 (0.22)	1.23	0.232
Happiness	4.59 (0.21)	3.79 (0.16)	4.27	<0.001**
Sadness	1.85 (0.14)	2.84 (0.18)	5.86	<0.001**

** $P < 0.001$.

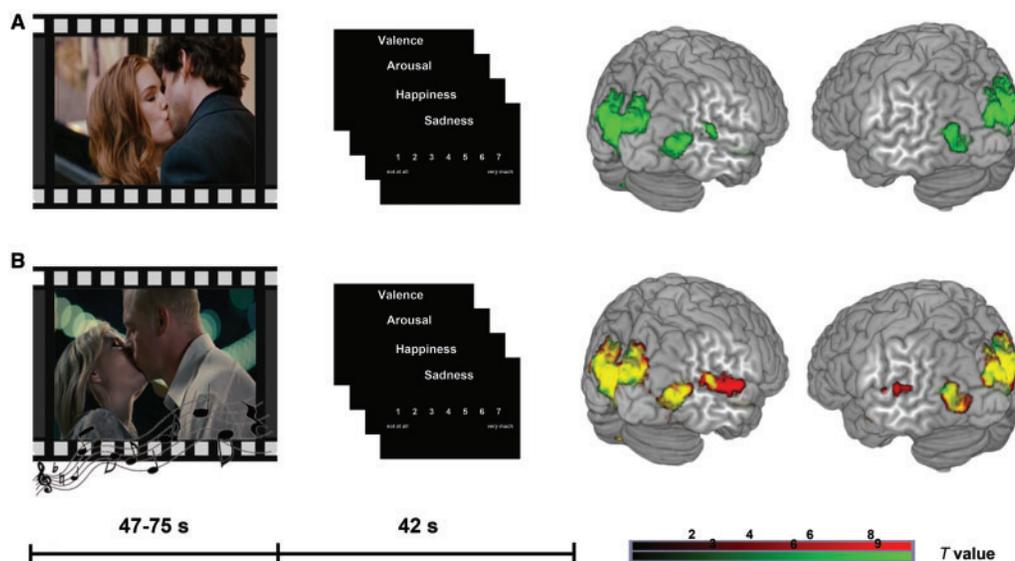


Fig. 1 Experimental design and brain activations. In the upper panel, the mute kissing scenes are illustrated with behavioral ratings and brain activations (A, green). The lower panel depicts both audiovisual conditions (B, red). The additive overlay is indicated by yellow colored activations (B, yellow). All activations depicted $P < 0.05$ FWE whole-brain corrected, cluster extent threshold $k > 5$ voxels.

fMRI experiment

With regard to stimulus combination and order of presentation, we employed a fully randomized design. For each participant, three conditions were randomly assigned to the 24 film clips resulting in eight film clips with happy music, eight film clips with sad music and eight film clips without music. The 16 music pieces were allocated randomly to the film clips to avoid a constant combination of one film clip with a specific piece of music. In a break of 42 s between the film clips, the participants rated their own emotional state in terms of valence, arousal, happiness and sadness. Participants gave their ratings using a trackball to move a cursor on a white scale (valence scale -3 to 3 ; arousal-, happiness- and sadness scale $1-7$) on a black background (Figure 1). For each item, they had to respond within 8 s.

Visual stimuli were presented using goggles (Resonance Technology, Northridge, CA, USA) and musical stimuli were presented binaurally through MRI-compatible headphones with standard comfortable volume of about 70 dB. All stimuli were presented using the stimulation software Presentation (Version 9.00, Neurobehavioral Systems, Albany, CA, USA). The whole fMRI experiment lasted 45.5 min and all data were acquired in a single session.

Functional MRI

The imaging data were acquired with a 3 T Siemens (Erlangen, Germany) Tim Trio MRI scanner using a 12-channel phased-array headcoil. After a high resolution T1-weighted structural image (TR 1900 ms; TE 2.52 ms; flipangle 9° ; voxelsize 1 mm^3 ; 176 sagittal slices; 1 mm slice) for registration of the functional data, a whole brain T2*-sensitive gradient-echo-planar-imaging sequence was applied (TR 2000 ms; TE 30 ms; isotropic voxelsize 3 mm^3 ; 1356 scans; flipangle 70° ; FOV $192 \times 192 \text{ mm}$; matrix 64×64 ; 37 slices; 3 mm slice thickness; 0.6 mm gap).

Behavioral data analysis

Behavioral data were analyzed with SPSS19 (Inc., 2009, Chicago, IL, www.spss.com). Two paired t -tests for the four emotion measures (valence, arousal, happiness and sadness) were calculated to compare

(i) the influence of the audiovisual conditions (happy and sad) to the mute condition and (ii) the differential effects of the two combined conditions (happy vs sad).

fMRI data analysis

The data were analyzed using SPM8 (Wellcome Department of Cognitive Neurology, London). Before preprocessing, the origin of the functional time series was set to the anterior commissure. Motion correction was performed using each subject's mean image as a reference for realignment. The corrected images were coregistered to the mean functional image generated during realignment. The T1 images were segmented using SPM8 (Ashburner and Friston, 2005) and the individual segmented grey and white matter images were used to create an average group MNI-registered template and individual flowfields using an image registration algorithm implemented in DARTEL tools (Diffeomorphic Anatomical Registration Through Exponentiated Lie Algebra; Ashburner, 2007). The realigned images were normalized to the final template using the previously generated flowfields. The images were resampled to a 2 mm isotropic voxelsize resolution and subsequently smoothed with a 6 mm full-width at half maximum isotropic Gaussian kernel. Statistical analysis was performed in the context of the general linear model (GLM) using a two-level approach (Friston et al., 1995). On the first level, the three conditions sad music, happy music and without music (mute condition) were modeled as block predictors for the BOLD activity over the whole length of the film clips. To account for movement-related variance, the realignment parameters were included as additional regressors in the design matrix. t -contrasts were computed for the mute condition, for both music conditions combined (happy and sad), as well as for the happy condition and the sad condition separately. On the second level, first a one-sample t -test was performed for the mute condition to test for brain regions involved in processing the dynamic visual stimuli. Two paired t -tests were performed to test for the effect of music (happy and sad music together) compared with the mute condition as well as each music condition separately to test for an emotion effect of music (happy music vs sad music). All results are reported at $P < 0.05$, family-wise error (FWE) corrected for multiple comparisons and a voxel extent threshold of 5. Based on a priori hypotheses, small

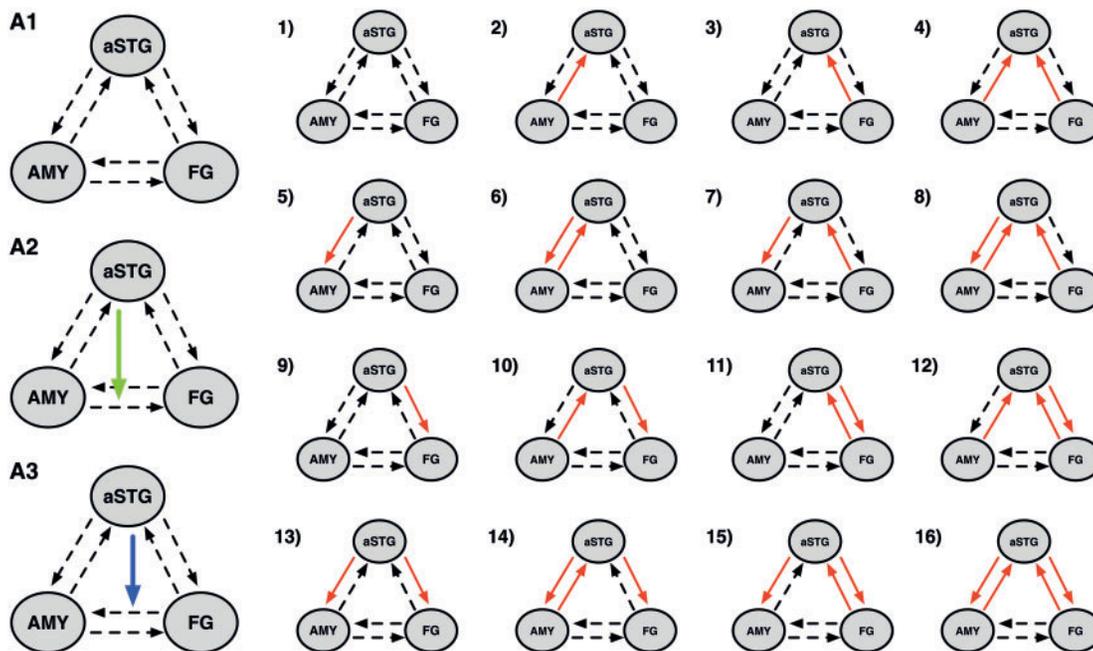


Fig. 2 Model space. Models 1–16 present all possible forward and backward connection and bidirectional modulations of happy and sad music as red arrows. (A1) Initial model family of connectivity without non-linear modulation of aSTG. (A2) Includes the non-linear modulation on the connection between AMY and FG. (A3) Winning family with a non-linear modulation on the connection from FG to AMY. All models were tested within the two non-linear families, except for model 1 in the bilinear family (A1), resulting in a model space of 47 models total. To note, the same model space was tested in the two hemispheres separately. For clarity the red arrow shows the modulatory input of music, but in fact happy and sad music were modeled separately.

volume corrections were performed for AMY and FG using anatomical masks taken from the automatic anatomic labeling atlas (Tzourio-Mazoyer *et al.*, 2002).

Dynamic causal modeling

Overview

DCM10 as implemented in SPM8 was used to analyze effective connectivity in an audio-visual-limbic (a three-region) brain network. This approach was chosen to study the neural dynamics underlying emotion processing during multisensory perception using movies combined with music. The emotional context was systematically manipulated by music. In the DCM framework, regional time series are used to analyze connectivity and especially its modulation by experimental conditions. The focus of DCM is the modeling of hidden neuronal dynamics to examine the influence that one neuronal system exerts over another (Friston *et al.*, 2003). Bilinear DCM models three parameters: (i) the impact of experimental stimuli can be modeled directly on specific regions (driving inputs); (ii) the endogenous coupling between two regions, which is context independent (intrinsic connections); and (iii) change of the strength of coupling between two regions driven by an experimental manipulation (modulatory input). This model was extended by a non-linear term that describes the influence of activity in one region on the coupling of two other brain regions (Stephan *et al.*, 2008). In DCM for fMRI, the modeled neuronal dynamics are transformed to a hemodynamic response by using a hemodynamic forward model (Friston *et al.*, 2000; Stephan *et al.*, 2007). Parameter estimation is performed in a Bayesian framework as described previously (Friston *et al.*, 2003).

Regions and time series extraction

Previous studies repeatedly reported AMY and FG to be involved in emotion processing during audiovisual integration (Dolan *et al.*, 2001; Jeong *et al.*, 2011; Müller *et al.*, 2011, 2012). Anterior and posterior

parts of the superior temporal sulcus were proposed to play an important role in audiovisual integration (Hocking and Price, 2008; Robins *et al.*, 2009; Müller *et al.*, 2011). The selection of these three regions was further supported by the present univariate results, which replicate the involvement of the STG in audiovisual integration and additionally showed an effect of happiness (happy>sad music) in aSTG. FG and AMY were both activated during the presentation of movies with and without music.

Regional time series were extracted for each hemisphere on the single-subject level using a combination of functional and anatomical criteria: second-level analysis clusters revealed by the *t*-contrast happy>sad music (for aSTG) and a *t*-contrast including all conditions (for FG and AMY) were masked with anatomical regions of interest (ROIs) (STG, FG, AMY) taken from the WFU Pick Atlas toolbox. Each subject’s MNI coordinates of the highest *t*-value (happy>sad music for aSTG, contrast including all conditions for FG and AMY) within these combined ROIs were surrounded with a sphere of 6 mm and used to extract the first eigenvariate (for mean coordinates of FG and AMY time series extraction; see Supplementary Table S7).

Model space

All models shared intrinsic connections between all of the three regions in both directions. All experimental blocks were defined as driving input to FG and only the combined conditions as driving input to aSTG. We hypothesized the presence of a gating effect of the aSTG on amygdalar–fusiform connectivity resulting in two possible non-linear modulations of aSTG on this connection (Figure 2, A2 and A3). This was theoretically and empirically motivated by work that underlines the role of the aSTG as heteromodal convergence zone mediating more complex modulatory influences (Calvert and Thesen, 2004; Robins *et al.*, 2009; Müller *et al.*, 2012).

Because aSTG was the only region showing (i) activation in response to music and (ii) differential effects with regard to emotion, bilinear

effects were only modeled for connections including the aSTG to elucidate how the emotional information is processed between FG and aSTG as well as AMY and aSTG. To explore the directionality of this effect, 16 models with bilinear modulations by happy and sad music including all possible combinations of the two music conditions on these four connections were constructed. These models were stratified in three families, two 'non-linear' families and one 'bilinear' family: the first two families included the original non-linear model, (i) aSTG on FG to AMY (Figure 2, A3) and (ii) aSTG on AMY to FG (Figure 2, A2) with the 16 bilinear models, resulting in 16 models in each of the two families (Figure 2). The bilinear family contained 15 bilinear models without a non-linear modulation (Figure 2, A1). This resulted in a total of 47 models separated in three model families. No inter-hemispheric connections were modeled and the same model space was estimated for both hemispheres separately. This was done to reveal a potential mechanism underlying the observed univariate results, in which no laterality effect was demonstrated. To note, this procedure is a simplification of a basic principle of brain function, i.e. the integration and coordination of activity from two hemispheres. This simplified approach was chosen to keep the complexity of the model space as low as possible.

Bayesian model selection

A comparison of the above-described model space was performed using Bayesian model selection (BMS) on two family levels in each hemisphere. This was done to reduce the model space systematically and to thereby examine the functional architecture of the network. First, we compared models with vs without a non-linear modulation. Due to the hypothesized gating effect, we assumed a better fit of models with a non-linear modulation. If so, we aimed to explore whether this gating effect by the aSTG takes place on the connection from FG to AMY or from AMY to FG. Finally, we looked at single models within the best-fitting family and asked how the emotional information is processed between the FG and aSTG as well as AMY and aSTG. For BMS, we used a novel random-effects approach for group studies (Stephan et al., 2009). BMS takes into account the fit of the models, primarily based on the number of free parameters, in relation to the model complexity. With higher complexity, the relative fit of a model may increase but generalizability may be reduced. Random-effects BMS gives so-called exceedance probabilities (EP), the probability that one model is more likely than another. The BMS version integrated in DCM10 for SPM8 also allows BMS on family levels to identify families of models, which are more likely than others (Penny et al., 2010). Bayesian model averaging (BMA) provides an average of parameters of an entire model space or a family of models weighted by the posterior probability of each model considered (Penny et al., 2010) and is a reasonable approach when no convincing EP of a single model is observed in a model family or when model fit differs in between-group studies (also compare Seghier et al., 2011; Deserno et al., 2012). For classical inference on connectivity parameters, *t*-tests were used and Bonferroni corrections for multiple comparisons were applied, if required.

RESULTS

Behavioral data

After each film clip the participants rated their own emotional state in terms of valence (scale -3 to 3), arousal, happiness and sadness (scale 1-7). To test for the effect of music, mean arousal and valence ratings of the happy and sad music were computed and compared with the mute

Table 2 Regional brain activation during the contrast happy vs sad

Region	Brodman area	x	y	z	Cluster size	t-value
Happy music vs sad music						
R. aSTG	BA 22	58	-12	4	32	10.43
L. aSTG	BA 22	-56	-14	4	22	10.11

Reported regions show a significant activation at $P < 0.05$ whole-brain FWE-corrected, cluster extent > 5 voxels.

condition using a paired *t*-test (Table 1). The valence and arousal ratings of the music conditions show no significant differences to the mute condition.

To test for the emotion effect of music, the ratings of sad stimuli were compared with those of happy stimuli using paired *t*-tests. The ratings show significant differences for the evoked sadness and happiness (both P 's < 0.001) but no significant differences for valence and arousal (Table 1).

fMRI-GLM

The processing of mute kissing scenes involved a brain network of primary and secondary visual cortices, including FG and brain regions associated with emotion processing as AMY, hippocampus and medial prefrontal cortex (FWE-corrected for the whole brain, $P < 0.05$, cluster extent > 5 voxels; Figure 1, Supplementary Table S3). These regions were also activated in both audiovisual conditions (Figure 1). The additional effect of music (music $>$ mute) revealed activation in primary and secondary auditory cortices and the left superior colliculus (Supplementary Table S4). With regard to the emotion effects of music, we compared happy and sad music conditions, and this contrast revealed a stronger activation in bilateral aSTG (happy $>$ sad, FWE-corrected for the whole brain, $P < 0.05$, cluster extent > 5 voxels, Table 2, Figure 3). No difference for sad $>$ happy was observed. Small volume corrections were applied (music $>$ mute, happy $>$ sad and sad $>$ happy) for AMY and FG, but no significant effect was detected (for correlations of experienced emotions with DCM parameter estimates; see Supplementary Analysis 1).

fMRI-DCM

For both hemispheres, BMS revealed that models with a non-linear modulation of aSTG on fusiform-amygdalar connectivity (32 models) clearly outperformed models with bilinear modulation only (family EP right 98.83%, EP left 70.30%; Figure 4). Within the 32 non-linear models, there still was convincing superiority for the family with a non-linear modulation of aSTG on FG to AMY connectivity (EP right 98.78%, EP left 67.09%; Figure 4). Looking for the best fitting model within the 16 models in this family, a very similar model structure in both hemispheres represented most likely Model 4 (Figure 2, model A3-4) with two additional bilinear modulations on the connections from FG to aSTG and on the connections from AMY to aSTG was identified as best fitting for the right hemisphere (EP 93.18%; Figure 4). For the left hemisphere, BMS revealed slightly more variable results. Models 4 (EP 45.71%, Figure 4) and 3 show a nearly equal fit. Model 3 (Figure 2, model A3-3) differs from model 4 only with regard to the absence of a bilinear modulation on AMY to aSTG and a slightly higher probability (EP 48.75%; Figure 4). Because no model had an outstanding EP value, connectivity parameters were assessed using BMA for the winning model family in the left hemisphere.

With regard to inference on model parameters, we focused on the experimental manipulations, namely the two bilinear modulations and the non-linear modulation for both hemispheres. The non-linear modulation of aSTG on the connection from FG to AMY was

significantly negative in the left hemisphere [$t(21)=2.17, P=0.04$; Table 3]. The same effect was observed for the right hemisphere where it approached significance [$t(21)=2.0, P=0.058$; Table 3].

Bilinear modulations of the most plausible right-hemisphere model (Figure 2, model A3-4) reached Bonferroni-corrected significance for

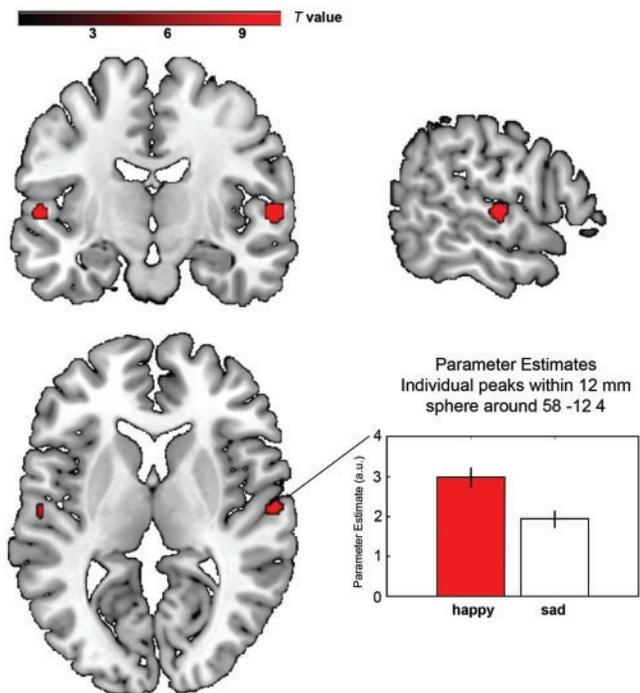


Fig. 3 Regional brain activations. The emotion effect of music (happy > sad) is located in the aSTG. The plot shows parameter estimates including betas at the individual peak within a 12 mm sphere around the group peak in the right aSTG with error bars indicating standard errors. All activations depicted $P < 0.05$ FWE whole-brain corrected, cluster extent threshold $k > 5$ voxels. Maximum t values in the left hemisphere $t = 10.11$ and in the right hemisphere $t = 10.43$. Coordinates are listed in Table 2.

sad music and happy music on the connection from FG on aSTG [happy $t(21)=2.94, P=0.031$; sad $t(21)=9.02, P < 0.001$; Table 3] and on the connection from AMY on aSTG [happy $t(21)=4.66, P < 0.001$; sad $t(21)=4.12, P=0.0019$; Table 3]. For the same model in the left hemisphere, bilinear modulations of sad music on the connection from FG on aSTG reached significance [Bonferroni-corrected, $t(21)=6.29, P < 0.001$; Table 3]. The same parameters reached significance when testing model 3 (Figure 2, model A3-3), ranking second based on BMS for the left hemisphere with no bilinear modulation on AMY to aSTG (Supplementary Table S5). All results remained significant when testing parameters derived from BMA (Supplementary

Table 3 DCM parameters. Mean connectivity parameters for model 4 (Figure 2, model A3-4) comprising bilinear modulation by music (happy and sad) and non-linear effects with standard errors (s.e.) in brackets. t -Tests were performed and t - and P -values are reported.

	Mean (s.e.)	t -value	P -value
Happy music			
L. FG \Rightarrow L. aSTG	-0.44 (0.35)	1.26	0.22
L. AMY \Rightarrow L. aSTG	-0.63 (0.44)	1.42	0.16
R. FG \Rightarrow R. aSTG	-0.78 (0.26)	2.94	0.031*
R. AMY \Rightarrow R. aSTG	-2.12 (0.45)	4.66	<0.001**
Sad music			
L. FG \Rightarrow L. aSTG	-1.60 (0.25)	6.29	<0.001**
L. AMY \Rightarrow L. aSTG	-0.46 (0.40)	1.15	0.26
R. FG \Rightarrow R. aSTG	-1.93 (0.21)	9.02	<0.001**
R. AMY \Rightarrow R. aSTG	-1.33 (0.32)	4.12	0.002*
Sad music > happy music			
L. FG \Rightarrow L. aSTG	1.15 (0.24)	4.66	<0.001**
L. AMY \Rightarrow L. aSTG	0.16 (0.37)	0.45	0.65
R. FG \Rightarrow R. aSTG	1.14 (0.20)	5.69	<0.001**
R. AMY \Rightarrow R. aSTG	0.79 (0.47)	1.65	0.11
Nonlinear modulation			
L. aSTG on L. FG \Rightarrow L. AMY	-0.64 (0.29)	2.17	0.041
R. aSTG on R. FG \Rightarrow R. AMY	-0.61 (0.30)	2.00	0.058

* $P < 0.05$; ** $P < 0.001$.

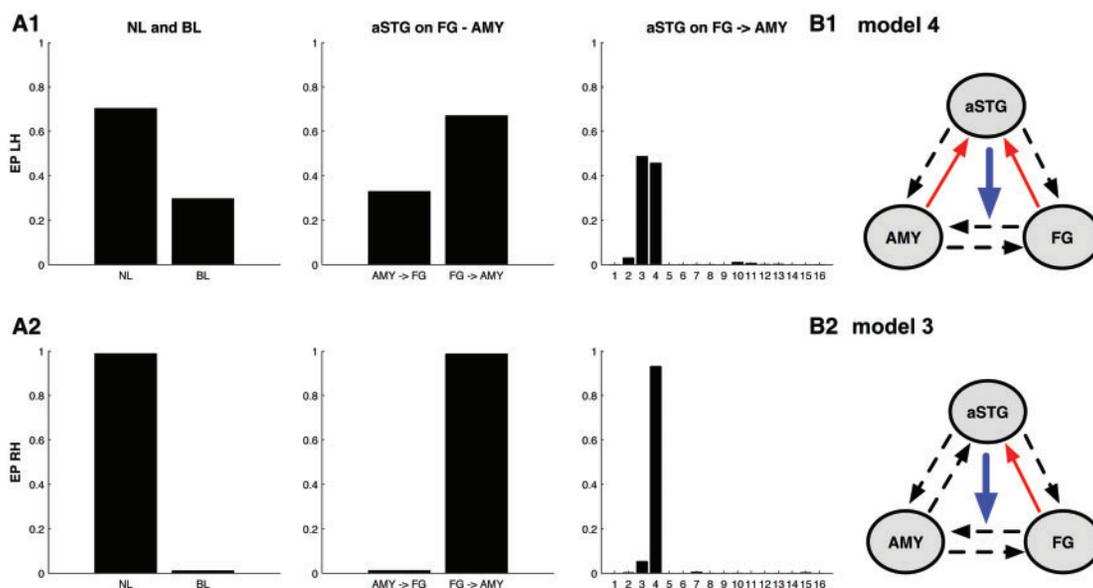


Fig. 4 Bayesian model selection. Results of Bayesian model selection for the left Hemisphere (LH, A1) and the right hemisphere (RH, A2). EP are reported as a measure of relative model fit. From left to right: Family selection of bilinear (BL) and non-linear (NL) model families, family selection of the two non-linear families: aSTG on AMY to FG connectivity (FG \Rightarrow AMY) and model selection for the model-subspace containing 16 models with all possible combinations of happy and sad music modulation. The winning models are number 3 (B2—model 3) and 4 (B1—model 4) for the left hemisphere and model 4 (B1—model 4) for the right hemisphere with backward projections from AMY and FG to aSTG.

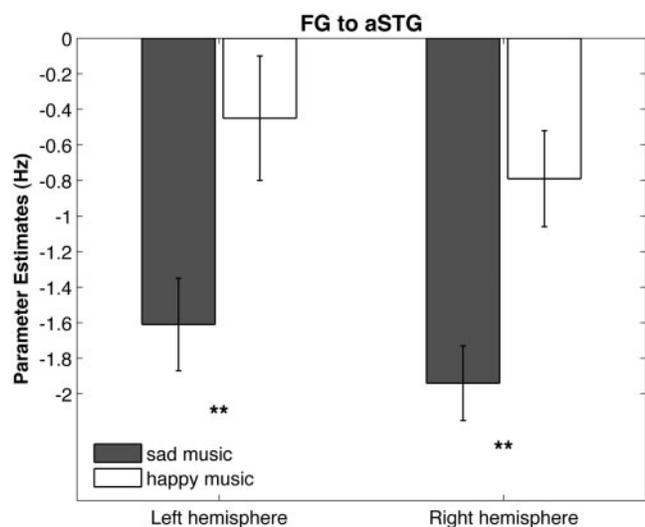


Fig. 5 Bilinear Modulation on the connection from FG to aSTG. In both hemispheres, suppressing influence of sad music on the connection from FG to aSTG was significantly stronger for sad than for happy music resulting in enhanced neural propagation of the FG to AMY. Error bars indicate standard error. $***P < 0.001$.

Table S8). All parameters for the modulation from FG to aSTG and for the modulation from AMY to aSTG were negative (Table 3). For comparison of happy and sad music, the parameters of the most plausible model (Figure 2, model A3-4) for the connection from FG to aSTG and from AMY to aSTG were compared using paired *t*-tests (Table 3). For both hemispheres, the parameters of sad music were significantly more negative than parameters of the happy music on the connection from FG to aSTG [right $t(21) = 5.69$, $P < 0.0001$; left $t(21) = 4.66$, $P < 0.001$; Table 3, Figure 5]. For the connection from AMY to aSTG, no significant differences were found (Table 3). Parameters for the intrinsic connectivity and driving input are provided in the Supplemental Data (Supplementary Table S6).

DISCUSSION

This study was designed to examine the modulatory effect of emotion-inducing music on the perception of dynamic kissing scenes. Bilinear and non-linear DCM was used to clarify the underlying mechanism of an emotion effect in the aSTG for happy compared with sad music. The results reveal emotion-specific activity in a functional network consisting of a superior temporal integration region, FG and AMY. Moreover, we show that the role of the aSTG in audiovisual integration of complex, naturalistic stimuli is that of a gating structure, controlling the connectivity between FG and AMY in an emotion-specific manner with more pronounced connectivity for sad compared with happy music underlying the kissing scenes. Bayesian model comparison revealed a winning model family that included a non-linear influence of aSTG on the connection from FG to AMY. The winning model within this family is characterized by additional bilinear modulation from FG and AMY to aSTG during the combined music–movie conditions. Thus, in line with our first hypothesis, the aSTG integrates emotional input from FG and AMY. All modulatory connectivity parameters were negative and therefore demonstrate suppression effects. Based on the results, the following mechanism may be proposed: the connection from FG to AMY is attenuated by activation of aSTG, whereas activation in FG and AMY in turn has a suppressing influence on aSTG. In accordance with our second hypothesis, the fusiform to amygdala coupling strength is gated via reduced suppression of the aSTG as region for multisensory integration of emotional material

Emotion effect in the aSTG

The emotion effect of this study is characterized by stronger BOLD signal changes in the bilateral aSTG for the happy music condition compared with the sad music condition. It is located in the anterior part of the STG and thus consistent with a region found to be involved in multisensory integration of emotional content (Robins et al., 2009). Because the pace of happy and sad music was identically balanced in this study, the result is best attributed to the differences in emotional quality (happiness and sadness) of the music.

The aSTG gates fusiform–amygdalar coupling

DCM enabled us to identify the aSTG as a gating structure that exerts a modulation on fusiform–amygdalar connectivity when emotional (happy or sad) music accompanies a kissing scene. Such second-order, indirect modulation through neurophysiological transmission via synaptic connections from the aSTG can occur directly on other connections or may also be neurobiological implemented by encompassing interneurons of adjacent neuronal populations.

Due to its suppressing impact, less activation in the aSTG causes enhanced information transfer from FG to AMY, as is particularly the case in the sad music condition. Dolan et al. (2001) proposed changes in fusiform–amygdalar connectivity to convey emotional effects in emotionally congruent conditions. Using effective connectivity we now can provide direct empirical evidence for this assumption by demonstrating increased forward connections from FG to AMY by co-presentation of emotional music. Enhanced fusiform–amygdalar coupling might be the basic principle for how a movie soundtrack modulates the emotional engagement of the spectator and the aSTG appears to play a major role in controlling this effect. A similar gating mechanism of the posterior superior temporal sulcus (pSTS) on fusiform–amygdalar connectivity was recently reported by Müller et al. (2012). They used happy, fearful and neutral faces combined with emotional sounds and conducted a DCM including auditory and visual cortices, the pSTS and the left amygdala. However, all their parameters were positive, whereas they are consistently negative in this study and hence reveal a suppressing effect of music on the couplings from FG and AMY to aSTG. The suppressive coupling is consistent with the univariate results, considering that neither FG nor AMY showed increased BOLD signal in response to the music. The absence of music-induced or emotion-specific activation in FG and AMY compared with the aSTG is also reflected by means of the endogenous connectivity parameters that are negative from FG to aSTG but positive between FG and AMY, which is in line with the observation that these two regions are similarly co-activated throughout all experimental conditions (see Supplementary Table S6).

Interestingly, the proposed mechanism still keeps the same effect as described by Müller et al. (2012): the more input is provided to the gating structure, the stronger is its strengthening effect on amygdalar–fusiform connectivity. The opposite directionality of the parameters could be due to the different stimulus material. We used more naturalistic input by combining film and music instead of combining static faces and sounds. Furthermore, DCM has undergone major revisions in recent years (Marreiros et al., 2008; Stephan et al., 2008; Daunizeau et al., 2009; Li et al., 2011). Müller and colleagues used an earlier version of DCM (DCM8), which may also contribute to differences in observed non-linear effects.

Although Müller et al. (2012) have found no context manipulation by emotion presented in both channels, our study provides strong evidence for the modulatory input of emotional music.

Soundtrack valences modulate the gating function of the aSTG

In addition to this unprecedented modulation of music, the models contained happy and sad music as distinct modulatory inputs, such

that we were able to distinguish the modulatory effects of these sound-track valences separately. The results revealed that sad music led to a stronger FG to aSTG suppression than happy music by reducing the aSTG gating and thereby yielding a strengthened fusiform–amygdalar coupling. This result additionally indicates that increased BOLD response in the aSTG during happy music might be explained by weaker suppressive influences from FG.

Relating to AMY, this valence effect may function to decrease stimulus ambiguity. As a potential self-relevancy detector (Sander *et al.*, 2003; Adolphs, 2010) the AMY seems more involved through higher vigilance based on stimulus ambiguity (Whalen, 1998). According to Whalen, ambiguous or less explicit stimuli require that the organism gathers more information to determine the appropriate behavior to engage in. Given that the kissing scenes are highly positive material, the combination with sad music may be perceived as more ambiguous and less congruent than with happy music.

LIMITATIONS

Due to a lack of cross-modal studies using naturalistic stimulus material, the ROIs chosen for the DCM analysis were derived mainly from multisensory integration studies, which used incongruity/congruity tasks with static faces and emotional sounds. However, in demonstrating connectivity changes between these regions using natural dynamic stimuli, this study strongly supports a participation of these regions when our brain forms a coherent percept from different senses, particularly when they provide emotional information. Yet, our results are based on highly positive film material, and would need to be replicated for generalizability to other, especially negative audiovisual material. Another constraint of this study is related to the organization of the human brain in two interconnected hemispheres. The present results show a more variable model fit in the left hemisphere calling future studies to consider inter-hemispheric connections during multisensory integration of emotion.

CONCLUSION AND OUTLOOK

This study shows that emotion elicitation through film takes place in a dynamic brain network that integrates multisensory emotional information, rather than in single functionally specialized brain areas. Robust evidence is provided that the existence of an emotion-inducing effect of music changes the interconnections between the nodes of a three-region network, consisting of aSTG, FG and AMY, thereby leading to a different emotional experience in the spectator. The aSTG was identified as a multisensory integration region for emotion evoked by music and as a structure that exerts control over the connection from the FG to the AMY. Comparison of DCM parameters revealed a significantly stronger suppressive influence on the connection from FG to aSTG during sad music than during happy music leading to less aSTG activation. As stronger aSTG activation results in enhanced suppression of intrinsic connectivity from FG to AMY, this indicates a strengthening of fusiform to amygdala coupling during sad music. Thus, this study suggests that the soundtrack has a modulatory, emotion-specific impact during movie perception by variations of fusiform–amygdalar connection using the aSTG as a multisensory control region. During movie perception, the brain integrates visual, auditory and semantic information over time. Future research should examine the modulatory effect of semantic cognition on emotion elicitation in movies.

SUPPLEMENTARY DATA

Supplementary data are available at SCAN online.

Conflicts of Interest

None declared.

REFERENCES

- Adolphs, R. (2010). What does the amygdala contribute to social cognition? *Annals of the New York Academy of Science*, 1191, 42–61.
- Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *NeuroImage*, 38(1), 95–113.
- Ashburner, J., Friston, K.J. (2005). Unified segmentation. *NeuroImage*, 26(3), 839–51.
- Baumgartner, T., Lutz, K., Schmidt, C.F., Jäncke, L. (2006). The emotional power of music: how music enhances the feeling of affective pictures. *Brain Research*, 1075(1), 151–64.
- Bernardi, L., Porta, C., Sleight, P. (2006). Cardiovascular, cerebrovascular, and respiratory changes induced by different types of music in musicians and non-musicians: the importance of silence. *Heart*, 92(4), 445–52.
- Boltz, M.G. (2005). The cognitive processing of film and musical soundtracks. *Memory & Cognition*, 32(7), 1194–205.
- 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.
- Cohen, A.J. (2001). Music as a source of emotion in film. In: Juslin, P.N., Sloboda, J.A., editors. *Music and Emotion: Theory and Research*. New York, NY: Oxford University Press, pp. 249–79.
- Damasio, A.R. (1989). Time-locked multiregional retroactivation: a systems-level proposal for the neural substrates of recall and recognition. *Cognition*, 33(1–2), 25–62.
- Daunizeau, J., Friston, K.J., Kiebel, S.J. (2009). Variational Bayesian identification and prediction of stochastic nonlinear dynamic causal models. *Physica D*, 238, 2089–218.
- de Gelder, B., Bertelson, P. (2003). Multisensory integration, perception and ecological validity. *Trends in Cognitive Sciences*, 7(10), 460–7.
- de Gelder, B., Vroomen, J. (2000). The perception of emotions by ear and by eye. *Cognition and Emotion*, 14(3), 289–312.
- Deserno, L., Sterzer, P., Wüstenberg, T., Heinz, A., Schlagenhaut, F. (2012). Reduced prefrontal-parietal effective connectivity and working memory deficits in schizophrenia. *Journal of Neuroscience*, 32(1), 12–20.
- Dolan, R.J., Morris, J.S., de Gelder, G. (2001). Crossmodal binding of fear in voice and face. *Proceedings of the National Academy of Sciences of the United States of America*, 98(17), 10006–10.
- Eldar, E., Ganor, O., Admon, R., Bleich, A., Hendler, T. (2007). Feeling the real world: limbic response to music depends on related content. *Cerebral Cortex*, 17(12), 2828–40.
- Ethofer, T., Anders, S., Erb, M., et al. (2006). Impact of voice on emotional judgment of faces: an event-related fMRI study. *Human Brain Mapping*, 27(9), 707–14.
- Friston, K.J., Harrison, L., Penny, W. (2003). Dynamic causal modelling. *NeuroImage*, 19(4), 1273–302.
- Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.P., Frith, C.D., Frackowiak, R.S.J. (1995). Statistical parametric maps in functional imaging: a general linear approach. *Human Brain Mapping*, 2(4), 189–210.
- Friston, K.J., Mechelli, A., Turner, R., Price, C.J. (2000). Nonlinear responses in fMRI: the Balloon model, Volterra kernels, and other hemodynamics. *NeuroImage*, 12(4), 466–77.
- Gross, J.J., Levenson, R.W. (1995). Emotion elicitation using films. *Cognition and Emotion*, 9(1), 87–108.
- Hocking, J., Price, C.J. (2008). The role of the posterior superior temporal sulcus in audiovisual processing. *Cerebral Cortex*, 18(10), 2439–49.
- Jeong, J.W., Diwadkar, V.A., Chugani, C.D., et al. (2011). Congruence of happy and sad emotion in music and faces modifies cortical audiovisual activation. *NeuroImage*, 54(4), 2973–82.
- Kappelhoff, H., Bakels, J.H. (2011). Das Zuschauergefühl—Möglichkeiten qualitativer Medienanalyse. *Zeitschrift für Medienwissenschaft*, 5(2), 78–95.
- Koelsch, S. (2010). Towards a neural basis of music-evoked emotions. *Trends in Cognitive Sciences*, 14(3), 131–7.
- 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), 1445–56.
- Kuchinke, L., Kappelhoff, H., Koelsch, S. (2013). Emotion and music in narrative films: a neuroscientific perspective. In: Tan, S.-L., Cohen, A., Lipscomb, S., Kendall, R., editors. *The Psychology of Music in Multimedia*. London: Oxford University Press.
- Li, B., Daunizeau, J., Stephan, K.E., Penny, W., Hu, D., Friston, K. (2011). Generalised filtering and stochastic DCM for fMRI. *NeuroImage*, 58(2), 442–57.
- Macaluso, E., Frith, C.D., Driver, J. (2000). Modulation of human visual cortex by cross-modal spatial attention. *Science*, 289(5482), 1206–8.
- Marreiros, A.C., Kiebel, S.J., Friston, K.J. (2008). Dynamic causal modelling for fMRI: a two-state model. *NeuroImage*, 39(1), 269–78.
- McGurk, H., McDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264(5588), 746–8.
- Mesulam, M.M. (1998). From sensation to cognition. *Brain*, 121(6), 1013–52.
- Müller, V.I., Cieslik, E.C., Turetsky, B.I., Eickhoff, S.B. (2012). Crossmodal interactions in audiovisual emotion processing. *NeuroImage*, 60(1), 553–61.

- Müller, V.I., Habel, U., Derntl, B., et al. (2011). Incongruence effects in crossmodal emotional integration. *NeuroImage*, 54(3), 2257–66.
- Penny, W.D., Stephan, K.E., Daunizeau, J., et al. (2010). Comparing families of dynamic causal models. *PLoS Computational Biology*, 6, e1000709.
- Pourtois, G., de Gelder, B., Bol, A., Crommelinck, M. (2005). Perception of facial expressions and voices and of their combination in the human brain. *Cortex*, 41(1), 49–59.
- Pourtois, G., de Gelder, B., Vroomen, J., Rossion, B., Crommelinck, M. (2000). The time-course of intermodal binding between seeing and hearing affective information. *Neuroreport*, 11(6), 1329–33.
- Robins, D.L., Hunyadi, E., Schultz, R.T. (2009). Superior temporal activation in response to dynamic audio-visual emotional cues. *Brain and Cognition*, 69(2), 269–78.
- Sander, D., Grafman, J., Zalla, T. (2003). The human amygdala: an evolved system for relevance detection. *Reviews in the Neurosciences*, 14(4), 303–16.
- Seghier, M.L., Josse, G., Leff, A.P., Price, C.J. (2011). Lateralization is predicted by reduced coupling from the left to right prefrontal cortex during semantic decisions on written words. *Cerebral Cortex*, 21(7), 1519–31.
- Stephan, K.E., Kasper, L., Harrison, L.M., et al. (2008). Nonlinear dynamic causal models for fMRI. *NeuroImage*, 42(2), 649–62.
- Stephan, K.E., Penny, W.D., Daunizeau, J., Moran, R.J., Friston, K.J. (2009). Bayesian model selection for group studies. *NeuroImage*, 46(4), 1004–17.
- Stephan, K.E., Weiskopf, N., Drysdale, P.M., Robinson, P.A., Friston, K.J. (2007). Comparing hemodynamic models with DCM. *NeuroImage*, 38(3), 387–401.
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etolot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, 15(1), 273–89.
- Vitouch, O. (2001). When your ear sets the stage: musical context effects in film perception. *Psychology of Music*, 29(1), 70–83.
- Whalen, P.J. (1998). Fear, vigilance and ambiguity: initial neuroimaging studies of the human amygdala. *Current Directions in Psychological Science*, 7(6), 177–87.