

Authenticity affects the recognition of emotions in speech: behavioral and fMRI evidence

Matthis Drolet · Ricarda I. Schubotz · Julia Fischer

Published online: 29 October 2011

© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract The aim of the present study was to determine how authenticity of emotion expression in speech modulates activity in the neuronal substrates involved in emotion recognition. Within an fMRI paradigm, participants judged either the authenticity (authentic or play acted) or emotional content (anger, fear, joy, or sadness) of recordings of spontaneous emotions and reenactments by professional actors. When contrasting between task types, active judgment of authenticity, more than active judgment of emotion, indicated potential involvement of the theory of mind (ToM) network (medial prefrontal cortex, temporoparietal cortex, retrosplenium) as well as areas involved in working memory and decision making (BA 47). Subsequently, trials with authentic recordings were contrasted with those of reenactments to determine the modulatory effects of authenticity. Authentic recordings were found to enhance activity in part of the ToM network (medial prefrontal cortex). This effect of authenticity suggests that individuals integrate recollections of their own experiences more for judgments involving authentic stimuli than for

those involving play-acted stimuli. The behavioral and functional results show that authenticity of emotional prosody is an important property influencing human responses to such stimuli, with implications for studies using play-acted emotions.

Keywords Prosody · Mentalizing · Episodic memory · Affect · Intonation · Theory of mind

Introduction

The ability to feign emotions is presumed to be uniquely human and is not only relevant to the ability to deceive others, but also forms the foundation for the dramatic arts. In turn, pretense (Cowie et al., 2001; Ekman, Friesen, & O'Sullivan, 1988), pretend play (Rakoczy & Tomasello, 2006), and social deception are all key topics of interest in theory of mind (ToM) research (Bänziger & Scherer, 2007; C. Frith & Frith, 2007; Hoekert, Vingerhoets, & Aleman, 2010; Premack & Woodruff, 1978). Importantly, although little is known about the potential impacts of acting on emotion expression and perception (Goldstein, 2009), a large portion of research into emotion perception makes use of stimulus material obtained from actors.

Studies on human emotion expression that employ acted behavior either focus on prototypical behavior or assume acting to be analogous to authentic emotional expression, both in terms of its production by the sender and its effect on the receiver (Belin, Fillion-Bilodeau, & Gosselin, 2008; Ethofer et al., 2006; Hietanen, Surakka, & Linnankoski, 1998; Laukka, 2005). However, findings on social deception (Adolphs, 2009; Ekman, Davidson, & Friesen, 1990; Frick, 1985; Grezes, Berthoz, & Passingham, 2006; Scherer, 1986) and,

Electronic supplementary material The online version of this article (doi:10.3758/s13415-011-0069-3) contains supplementary material, which is available to authorized users.

M. Drolet · J. Fischer
German Primate Center,
Göttingen, Germany

M. Drolet · R. I. Schubotz
Max Planck Institute for Neurological Research,
Köln, Germany

M. Drolet (✉)
Cognitive Ethology Department, DPZ (German Primate Center),
Kellnerweg 4,
37077 Göttingen, Germany
e-mail: mdrolet@dpz.eu

recently, feigning emotions (Scheiner & Fischer, 2011; Wilting, Kraemer, & Swerts, 2006) indicate that acted behaviors may differ from more naturally occurring ones. Scheiner and Fischer collected 80 noninstructed “authentic” speech recordings of four emotional expression categories (20 each of anger, sadness, joy, and fear) from German radio archives and had each reenacted by a professional actor. Recordings were presented to participants from three separate cultures (Germany, Romania, Indonesia) in forced choice rating tasks of emotional content and recording condition (“authentic” vs. “play-acted”). Although participants were poor at distinguishing between conditions (only slightly above chance), these nevertheless influenced recognition of emotional content: Across all three cultures, anger was recognized more accurately when play acted, and sadness when authentic.

The principal goal of the present study was to determine what specific neuronal substrates are active during judgments on emotion and authenticity and, in a second step, whether stimulus authenticity additionally modulates the activation during these tasks. Participants were presented with emotion- and authenticity-judgment tasks in an fMRI setup to determine the activation loci involved during these explicit tasks. Contrasts based on stimulus authenticity could then be applied to determine its effect on activation during either task. As in the study by Scheiner and Fischer (2011), those expressions that were obtained during radio interviews are defined as noninstructed and were labeled as “authentic,” whereas acted emotions were produced by professional actors following scripted instructions. Although radio interviews may not exactly mirror more private expressions of emotion, these nevertheless represent unscripted emotional expressions in a nonlaboratory setting.

Explicit judgment of authenticity was predicted to activate the so-called ToM network. The ToM network is involved in the perception of oneself and other individuals as cognitive entities and, more specifically, in the determination of mental states (Bolton, 1902; Povinelli & Preuss, 1995; Premack & Woodruff, 1978). U. Frith and Frith (2003) reviewed a series of neuroimaging studies indicating that the medial prefrontal cortex (PFC), the temporal poles, and the posterior superior temporal sulcus (STS) (extending into the temporoparietal junction [TPJ]) are components of the ToM network. Van Overwalle and Baetens (2009) additionally incorporated the posterior cingulate (PCC) extending into the precuneus (PC) as part of a review of cognitive neuroscience research on mentalizing.

In addition to the ToM network, the stimuli were likely to activate regions of the brain involved in the processing of emotion and prosody. In this study, explicit judgments of emotional content versus a word detection control condition were expected to activate loci in the orbital and inferior frontal, superior temporal, and inferior parietal cortices (as

reviewed by Schirmer & Kotz, 2006). In addition, differential activation to individual emotions has been described in the cingulate cortex (Bach et al., 2008; Buchanan et al., 2000; Wildgruber et al., 2004; Wildgruber et al., 2005).

On the basis of the previous results by Scheiner and Fischer (2011), we predicted that, on a behavioral level, anger would be detected correctly most often when play acted, and sadness when authentic. Authentic and play-acted recordings were predicted to each have detection rates slightly above chance. Accordingly, stimulus authenticity (as defined previously) was expected to modulate activity in both prosody-specific regions and the ToM network so that play-acted emotions lead to higher activations than authentic emotions (bottom-up/stimulus-driven effects). This prediction was made on the assumption that actors are aware of the dichotomy between the emotion to be communicated and their current environment and state. This, if encoded in the stimuli, potentially represents additional information encoded in play-acted as opposed to authentic recordings.

Method

Participants

Neuroanatomical and performance effects on behavioral tasks can differ as a function of gender (Everhart, Demaree, & Shipley, 2006; R. C. Gur, Gunning-Dixon, Bilker, & Gur, 2002). Specifically, emotion recognition accuracy and reaction time (RT) show a distinct advantage for women (Belin et al., 2008; Hall & Schmid Mast, 2007; McClure, 2000; Schirmer, Kotz, & Friederici, 2005). Moreover, differences in the size and shape of the cingulate cortex have been found between men and women (Kochunov et al., 2005). Therefore, 24 female participants who were 20–30 years of age ($M = 24$ years of age; right-handed; German as mother-tongue) were selected and contacted using the Cologne MPI database for fMRI experiments. Only individuals without a history of neurological or psychological complications (including the use of psychiatric medication) were included. Participants were informed about the potential risks of magnetic resonance imaging and were screened by a physician. They gave informed consent before participating and were paid afterward. The experimental standards were approved by the local ethics committee, and data were handled pseudonymously.

Stimulus selection

Authentic recordings were selected from German radio archives. The recordings (mono wave format; sample rate of 44.1 kHz) were selected from interviews with individuals

talking in an emotional fashion (fear, anger, joy, or sadness) about a highly charged ongoing or recollected event (e.g., parents speaking about the death of their children, people winning in a lottery, in rage about injustice, or threatened by a current danger). Emotion was ascertained through the verbal content of the spoken text, as well as through the context either described in the recording details or in the recordings themselves. Although the potential influence of social acting can never be completely excluded, this effect was minimized by excluding clearly staged settings (e.g., talk shows) and relying on nonscripted interviews. Of the 80 speech tokens, 35 were made outdoors and varied in their noise surroundings, but all selected recordings were of good quality with minimal background noise and did not contain any keywords that could allow inference of the expressed emotion. To ensure inference-free verbal content, text-only transcripts were rated by participants who were naive to the purpose of the experiment. Recordings for which the respective emotion was recognized better than chance were replaced. The final original stimulus set consisted of 20 samples per emotion (half from male and half from female speakers), resulting in a total of 80 recordings made by 78 different speakers.

Play-acted stimuli were performed by actors from Berlin, Hanover, and Göttingen, in Germany (each replicated a maximum of three recordings of equivalent emotional content; total of 42 actors). The actors were told to express the respective text and emotion in their own way, using only the text, identified context, and emotion (the segment to be used as stimulus was not indicated, and the actors never heard the original recording). To remove category effects between authentic and play-acted stimuli, the environment for the play-acted recordings was varied by recording in different locations (30 recorded outdoors and 50 recorded indoors) while avoiding excessive background noise. Segments selected as stimuli (mean length = $1.76 \text{ s} \pm 1.01 \text{ SD}$; range = $0.357 - 4.843 \text{ s}$) did not contain any keywords that could allow inference of the expressed emotion. The average amplitude of all stimuli was equalized with Avisoft SASLab Pro Recorder Version 4.40 (Avisoft Bioacoustics, Berlin, Germany).

Trial and stimulus presentation

The program NBS Presentation (Neurobehavioral Systems, Inc., Albany, California) controlled the details of each experimental run with trial structure, timing, and order preprogrammed via scripts. Each experimental run (one per participant) included 178 trials (limited only by the number of stimuli available), of which 72 were used for an emotion judgment task and 72 for an authenticity judgment task. In addition, two control tasks were included: 16 word detection trials (utilizing the respective eight authentic and

eight play-acted stimuli not used in the experimental task trials), in which participants had to count occurrences of the word *und* (“and”), and 18 empty trials with pink-noise playback. The word detection task was included as additional control due to interest in the ToM network in this experiment. As has been shown previously (Buckner, Andrews-Hanna, & Schacter, 2008) the so-called default-mode network that is activated during rest trials partly mirrors the ToM network. In order to detect potential ToM activation, we therefore applied this additional control task as opposed to the empty trials, which could mask actual ToM activity if contrasted with the experimental trials (Spreng, Mar, & Kim, 2009). For emotion judgments, four responses were possible: anger, sadness, happiness, fear (presented in German as: *Wut, Trauer, Freude, Angst*), whereas for authenticity, judgment responses were presented as authentic (*echt*) and play acted (*theater*; described to participants beforehand as *gespielt*, i.e., play acted). To minimize eye movement, the maximal line-of-sight angle for visual information was kept under 5 degrees. Trial type and stimulus type pseudorandomizations were performed using conan (UNIX shell script: Max-Planck Institute for Neurology in Leipzig, Germany) to reduce any systematic effects that could have otherwise occurred with simple randomization. Each participant was shown a button sequence onscreen (800 x 600 pixel video goggles: NordicNeuroLab, Bergen, Norway) that was complementary to the response box layout (10 x 15 x 5 cm gray plastic box with a row of four black plastic buttons). For both emotion judgment and word detection, all buttons were assigned a possible response. For authenticity judgment only the two left-most buttons were used (Figs. 1, 2).

Experimental procedure

Participants were fitted with headphones for audio playbacks (NNL: NordicNeuroLab, Bergen, Norway) after they were placed in a supine position on the fMRI table. Imaging was performed with a 3T Siemens MAGNETOM TrioTim (Cologne, Germany) system equipped with a standard birdcage head coil. Participants were placed with their four fingers (excluding thumb) positioned on the

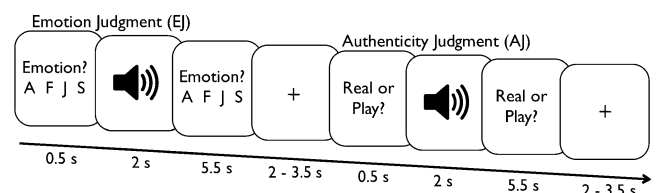


Fig. 1 Experimental trial sequences for emotion and authenticity judgment trials with durations in seconds (in seconds). A F J S represent the four possible emotion responses (forced choice design: anger, fear, joy, and sadness, respectively)

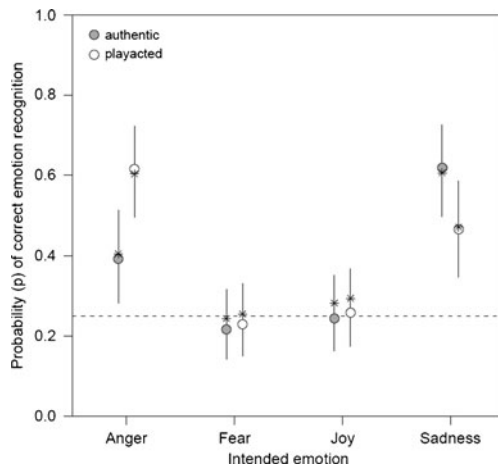


Fig. 2 Behavioral results from fMRI for emotion judgments. Actual emotion of recordings on *x*-axis, probability of recognition of emotion on left *y*-axis, logit function from statistical model on right *y*-axis. Circles indicate values calculated by model with 95% confidence intervals. Asterisks indicate actual mean recognition values

response buttons of the response box. Form-fitting cushions were utilized to prevent head, arm, and hand movements. Twenty-two axial slices (210-mm field of view; 64 x 64 pixel matrix; 4-mm thickness; 1-mm spacing; in-plane resolution of 3 x 3 mm) parallel to the bicommissural line (AC–PC) and covering the whole brain were acquired using a single-shot gradient EPI sequence (2,000-ms repetition time; 30-ms echo time; 90° flip angle; 1.8-kHz acquisition bandwidth) sensitive to BOLD contrast. In addition to functional imaging, 22 anatomical T1-weighted MDEFT images (Norris, 2000; Ugurbil et al., 1993) were acquired. In a separate session, high-resolution whole-brain images were acquired from each participant to improve the localization of activation foci using a T1-weighted 3-D-segmented MDEFT sequence covering the whole brain. Functional data were mapped onto this 3-D average using the 2-D anatomical images made immediately following the experiment. Including a visual and auditory test prior to the experiment, one experimental run lasted approximately 45 min.

Behavioral statistical analysis

Recognition accuracy was analyzed using the *R* Statistical Package v2.8 (R Development Core Team, 2008). The generalized linear mixed model (binomial family function) was implemented to determine the best model fit for recognition rates using the *glmer* function from the *lme4* package. The Akaike information criterion was used to select the model that best fit the data. Choice theory (Agresti, 2007) was implemented as a baseline-category logit model using the *multinom* function from the *VR* package (Smith, 1982). Model predictions are presented in figures using the logit values calculated by

R. Respective probability of correct responding is indicated. RTs \pm standard deviations in seconds (RTs are compared using a standard ANOVA) are also stated. The statistical analyses performed on the data showed main effects of and interactions between emotion and authenticity while accounting for individual participant effects. For the authenticity task, the d' value from signal detection theory was calculated from the recognition probability estimates as a measure of signal discrimination [$d' = z(\text{correct assessments}) - z(\text{false alarms})$].

fMRI statistical analysis

The functional data were processed using the software package LIPSIA v1.6.0 (Lohmann et al., 2001). This software package is available under the GNU General Public License (<http://www.cbs.mpg.de/institute/>). Functional data were motion corrected offline with the Siemens motion correction protocol. To correct for temporal offset between the slices acquired in one image, a cubic-spline interpolation was employed. Low-frequency signal changes and baseline drifts were removed using a temporal high-pass filter with a cut-off frequency of 1/95 Hz. Spatial smoothing was performed with a Gaussian filter of 5.65-mm FWHM. To align the functional data slices with a 3-D stereotactic coordinate reference system, a rigid linear registration with six degrees of freedom (three rotational, three translational) was performed. The rotational and translational parameters were acquired on the basis of the MDEFT slices to achieve an optimal match between these slices and the individual 3-D reference data set. The MDEFT volume dataset with 160 slices and 1-mm slice thickness was standardized to the Talairach stereotactic space (Talairach & Tournoux, 1988). The rotational and translational parameters were subsequently transformed by linear scaling to a standard size. The resulting parameters were then used to transform the functional slices using trilinear interpolation, so that the resulting functional slices were aligned with the stereotactic coordinate system, thus generating output data with a spatial resolution of 3 x 3 x 3 mm (27 mm³).

Statistical evaluation was based on a least-squares estimation using the general linear model for serially autocorrelated observations (Friston et al., 1998; Worsley & Friston, 1995). The design matrix was generated with a gamma function, convolved with the hemodynamic response function. Brain activations were analyzed time locked to recording onset, and the analyzed epoch was individually set for each trial to the duration of the respective stimulus (mean = 1.75 s; range = 0.35 – 4.84 s). The model equation, including the observation data, design matrix, and error term, was convolved with a Gaussian kernel of dispersion of 5.65-s FWHM to account for temporal

autocorrelation (Worsley & Friston, 1995). In the following, contrast images (i.e., beta value estimates of the raw-score differences between specified conditions) were generated for each participant. Since all individual functional data sets were aligned to the same stereotactic reference space, the single-subject contrast images were entered into a second-level random effects analysis for each of the contrasts. One-sample *t* tests were employed for the group analyses across the contrast images of all participants that indicated whether observed differences between conditions were significantly distinct from zero. The *t* values were subsequently transformed into *z*-scores. To protect against false-positive activation in task-based contrasts, the results were corrected for multiple comparisons by the use of cluster-size and cluster-value thresholds obtained by Monte Carlo simulations with a significance level of $p < .05$ (Lohmann et al. 2001). To protect against false-positive activation in stimulus-based contrasts, a voxel-wise *z*-threshold was set to $Z = 2.33$ ($p = .01$), with a minimum activation area of 270 mm^3 . Small volume correction was performed, correcting the results for a restricted search volume using spheres with a radius of 20 voxels ($p < .05$). The volumes of the individual spheres, each centered at a previously published coordinate (mPFC, -1 -56 33; TPJ, -51 -60 26, and 54 -49 22; retrosplenium, -3 51 20) according to the review by van Overwalle and Baetens (2009; Wurm, von Cramon, & Schubotz, 2011), were summed to calculate the alpha level of each individual activation cluster within the a priori hypothesized regions of the ToM network, based on the published coordinates reviewed by van Overwalle and Baetens (2009).

Results

Behavioral results

Recognition rates and significant differences as calculated via the generalized linear mixed model are presented in Table 1. Recognition rates for the emotion task produced significant interaction effects with authenticity of stimulus (Fig. 1; Supplement 1). Anger was detected correctly more often when play acted than when authentic, whereas sadness was detected correctly more often when authentic than when play acted. Fear was detected only near chance levels for both authentic and play-acted recordings. Recognition of joy was at chance levels for both authentic and play-acted recordings. Participants appear to have detected joy less well during the fMRI experiment than as reported by Scheiner and Fischer (2011). Because the model for authenticity recognition with emotion category as a factor did not produce any significant effects (Supplement 2), results are presented without emotion as a factor. Recogni-

Table 1 Recognition accuracy for emotion and authenticity trials

Recording Type		Emotion Trials		Authenticity Trials	
		PE	SE	PE	SE
Authentic	Anger	0.39	0.04 **		
	Fear	0.23	0.04 ***	0.70	0.05 ***
	Joy	0.27	0.04 **		
	Sadness	0.60	0.05 ***		
Play-Acted	Anger	0.60	0.05 ***		
	Fear	0.24	0.06 **	0.64	0.03 *
	Joy	0.28	0.07 **		
	Sadness	0.45	0.07 ***		

Probability estimates (PE) calculated from the generalized linear mixed model with standard errors (SE). Emotion trials required a determination of the emotional content, and authenticity trials a determination of the authenticity content of the recordings. Values for authenticity trials were calculated without effect of recording emotion since this produced a better fit for the model. * $p < .05$ ** $p < .01$ *** $p < .001$

tion was greater for authentic than for play-acted recordings, but the RTs did not differ significantly (RT authentic stimuli: 2.34 ± 0.39 s vs. play-acted stimuli: 2.41 ± 0.46 s), indicating that this was not due to general task difficulty. These recognition rates are above chance levels and indicate that authentic recordings were detected correctly more often than play-acted recordings (Fig. 3).

An ANOVA of RTs (Table 2) did not indicate any significant main effects of stimulus emotion ($F = 2.27$, $p > .05$) or of stimulus authenticity ($F = 3.33$, $p > .05$), or interactions ($F = 1.38$, $p > .05$) between emotion and authenticity. It did indicate a main effect for task type RTs ($F = 9.58$, $p < .001$), with authenticity task (RT = 2.38 ± 0.41 s) being faster than emotion task (RT = 3.24 ± 0.50 s). Because of the significant difference between RTs in the two experimental tasks, all contrast designs have each trial weighted by RT to ensure that any significant effects are not due to differences in task difficulty.

Using the probability estimate for correct authenticity detection as “hits” and incorrect responses of authentic for play-acted trials as “false alarms” (signal detection theory analysis) resulted in a d' value of 0.62 (indicates difficult discrimination between stimuli), as well as a decision criterion of -0.08 (very slight bias for play-acted responses). The choice theory analysis indicated that participants had a significant bias in assigning play-acted stimuli to anger and assigning authentic stimuli to sadness, and a significant bias against selecting joy (Fig. 4).

fMRI results

As an additional control for the effect of task difficulty in this contrast, the contrast of authenticity versus emotion

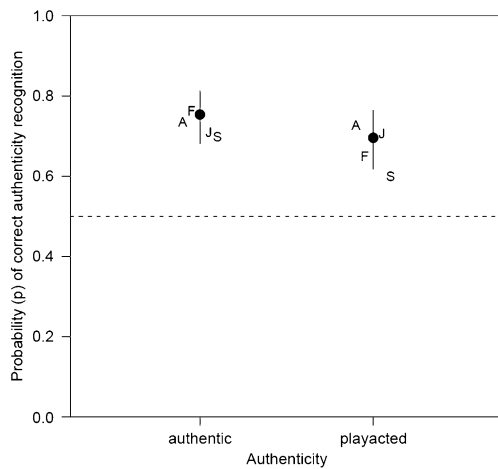


Fig. 3 Behavioral results from fMRI for authenticity judgments. Actual authenticity of recordings on x-axis, probability of recognition of authenticity on left y-axis, logit function from statistical model on right y-axis. Circles indicate values calculated by model with 95% confidence intervals. Letters indicate actual mean recognition values by emotion. A anger; F fear; H happy; S sad

judgment included RT as a random factor (regressor of no interest). Each trial was coded with the actual RT recorded for that trial. Activation loci with significantly greater activation for the authenticity judgment [conjunction (authenticity > emotion) ∩ (authenticity > word-detection)] included right inferior frontal gyrus (BA 47), right frontoparietal operculum, bilateral hippocampus, bilateral retrosplenial cortex, bilateral dorsal superior frontal gyrus, medial PFC, bilateral ventromedial PFC, bilateral temporoparietal junction, left superior temporal gyrus, left anterior cingulate cortex, and left anterior middle temporal gyrus (Fig. 5, Table 3). However, task weightings cannot completely eliminate a potential confounding influence of task difficulty.

Instances of participants listening to authentic and play-acted recordings were contrasted to detect stimulus effects due to stimulus authenticity. To ensure that significant modulation of resonance is not influenced by variable numbers of speakers between authentic and play-acted stimuli, the

Table 2 Reaction times (RTs) and standard errors (SEs) by stimulus type

	Authentic		Play Acted	
	RT(s)	SE	RT(s)	SE
Anger	3.25	0.60	2.65	0.53
Fear	3.22	0.63	3.19	0.50
Joy	3.43	0.70	3.23	0.59
Sadness	3.31	0.48	3.58	0.54

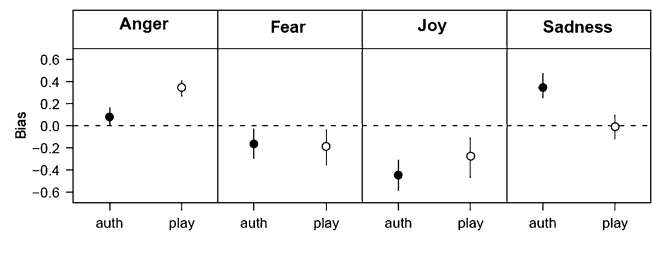


Fig. 4 Analysis of emotion recognition data in terms of the choice theory. Given is the response bias for each of the four possible choices (anger, fear, joy, sadness) dependent on authenticity. Data are given as means ± 95% confidence interval. In the absence of any bias, all four log-transformed bias values would be zero. Positive values indicate a bias toward choosing the response named in the headline; a value below zero indicates a bias against choosing the respective response. *auth* authentic; *play* play acted

contrast of authentic versus play acted included speaker repetition as a random factor (regressor of no interest). Repetition was defined for each participant separately, so that the first play-back by a specific speaker was labeled as “1,” the second as “2,” and the third as “3” (as appropriate for specific stimuli). This contrast was first performed for authenticity-judgment tasks to determine whether the network found previously is influenced by stimulus authenticity. This revealed significant activation loci in right medial PFC, left retrosplenium, left angular gyrus, and left dorsal superior frontal gyrus (Fig. 6; Table 4).

The determination of whether stimulus authenticity has an effect on emotion recognition was performed with the

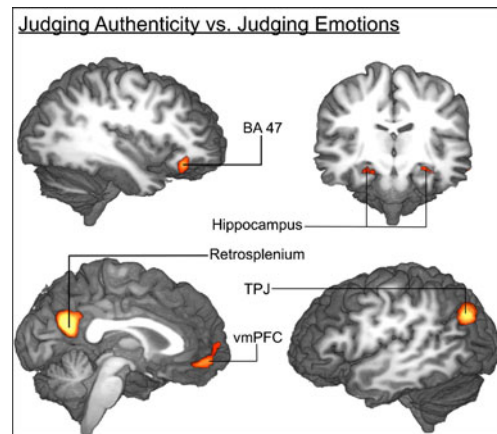


Fig. 5 Brain activation correlates of experimental tasks. Group-averaged (n = 24) statistical maps of significantly activated areas for authenticity versus emotion judgment trials. To correct for false-positive results, a voxel-wise z-threshold was set to Z = 2.33 (p = .01) with a minimum activation area of 270 mm³ and was mapped onto the best average subject 3-D anatomical map. Top left: right view sagittal section through temporal lobe. Top right: anterior view coronal section through thalamus. Bottom left: right view through corpus callosum. Bottom right: left view sagittal section through temporal cortex. *TPJ* temporoparietal junction; *vmPFC* ventromedial prefrontal cortex

Table 3 Network for judging authenticity [(Authenticity judgment > Emotion judgment) ∩ (Authenticity judgment > Word detection)]

Area	x	y	z	Z
Medial PFC	1	57	9	3.26
	-2	57	9	3.20
Ventromedial PFC	-8	15	-15	3.08
	4	33	-9	3.12
Hippocampus	-29	-36	-9	4.03
	-29	-18	-12	3.24
	19	-33	-12	3.35
	16	-17	-15	3.02
Ventral IFG	31	30	-9	4.97
Retrosplenium	-8	-60	21	5.76
	2	-57	24	5.21
TPJ	-44	-72	36	5.07

Anatomical specification, Talairach coordinates and maximum Z value of local maxima. *PFC* prefrontal cortex; *TPJ* temporoparietal junction; *STG* superior temporal gyrus; *SFG* superior frontal gyrus; *IFG* inferior frontal gyrus; *MTG* middle temporal gyrus

same authentic versus play-acted contrast but for emotion judgments, which revealed similar effects of authenticity on the activation seen in right medial PFC and left retrosplenium. Considering the average RTs that were measured from participant task completion and the similar recognition rates for authentic and play-acted stimuli, it is unlikely that any differences in activation were due to variability in participant concentration level or stimulus difficulty. Additional modulation detected by this contrast included left hemisphere activation of the middle frontal gyrus, precentral gyrus, superior parietal lobule, and middle occipital gyrus. (Fig. 6, Table 4).

Table 4 Direct effects of stimulus authenticity (authentic > play-acted)

Area	x	y	z	Z
During AJ				
Dorsal SFG	-23	24	45	3.43
Medial PFC	4	54	18	3.03
Ventromedial PFC	-8	36	-9	3.26
	1	36	-9	3.46
Ventral IFG	-41	27	-6	3.35
Anterior MTG	-62	-12	-12	2.81
Retrosplenium	-11	-54	12	3.17
TPJ	-47	-69	36	3.19
During EJ				
Medial PFC	1	51	15	2.88
MFG	-23	-15	57	3.51
PrG	-32	-27	51	3.72
SPL	-26	-54	57	2.87
Retrosplenium	-17	-48	9	2.85
	-5	-57	27	2.65
MOcG	-38	-90	0	3.84

Anatomical specification, Talairach coordinates and maximum Z value of local maxima ($p < .05$, corrected). *SFG* superior frontal gyrus; *PFC* prefrontal cortex; *IFG* inferior frontal cortex; *MTG* middle temporal gyrus; *TPJ* temporoparietal junction; *MFG* middle frontal gyrus; *PrG* precentral gyrus; *SPL* superior parietal lobule; *MOcG* middle occipital gyrus

Discussion

The present results demonstrate that the authenticity of emotional prosody has an effect on emotion recognition and modulates cerebral activation as reflected by BOLD fMRI.

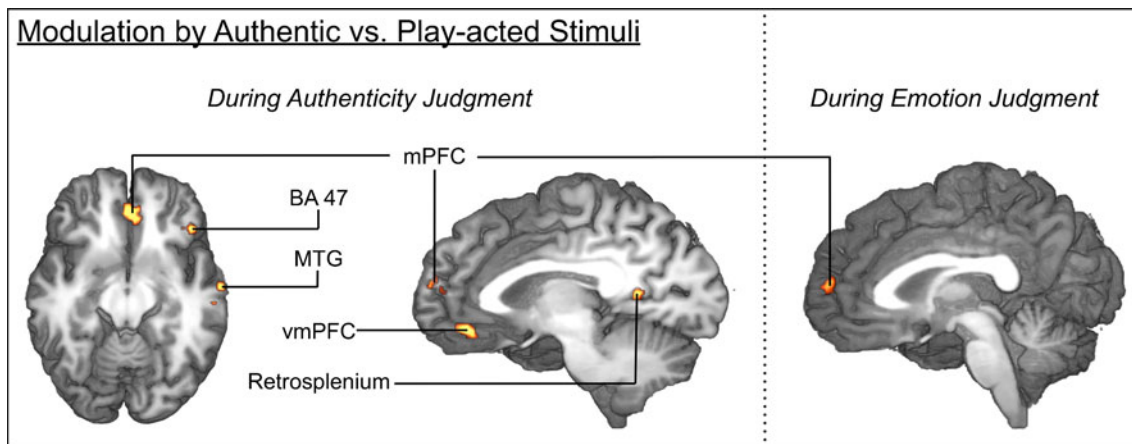


Fig. 6 Brain activation correlates of experimental tasks. Group-averaged ($n = 24$) statistical maps of significantly activated areas for authentic versus play-acted stimuli (authenticity judgments: left and middle; emotion judgments: right). To correct for false-positive results, a voxel-wise z -threshold was set to $Z = 2.33$ ($p = .01$) with a minimum activation area of 270 mm^3 and was mapped onto the best

average subject 3-D anatomical map. Left: ventral view axial section through temporal lobes. Middle: left view sagittal section through corpus callosum. Right: left view sagittal section through corpus callosum. *MTG* middle temporal gyrus; *vmPFC* ventromedial prefrontal cortex; *mPFC* medial prefrontal cortex

Explicit judgments about the authenticity of a stimulus, as compared with judgments about the emotional valence, may involve parts of the so-called ToM network. Activation was further boosted for authentic as compared with play-acted stimuli, as was shown for both judgments on emotion and on authenticity, which points to a significant impact of stimulus authenticity during the processing of emotional prosody, even when authenticity is not attended to explicitly. This task-independent modulation included the medial prefrontal cortex, a region of the ToM network known to be activated by stimuli with social context and in tasks requiring multimodal integration (Krueger, Barbey, & Grafman, 2009; Van Overwalle, 2009), as well as the retrosplenial cortex.

Authenticity judgments were contrasted with emotion judgments and a word detection control condition to ensure that any significant activation was indeed specific to the processing of authenticity. This contrast indicated an increase in activity for authenticity judgments in the medial PFC, the retrosplenium, and the right and left TPJ. Medial PFC and TPJ subserve a variety of functions linked to ToM and social cognition (Abraham, Werning, Rakoczy, von Cramon, & Schubotz, 2008; U. Frith & Frith, 2003) (Gallagher & Frith, 2003), communicative intent (C. Frith & Frith, 2007) and self-referential thought (Ochsner & Gross, 2005). TPJ activation points more specifically toward the inclusion of perspective taking, seen in both spatial perspective taking (Gallagher & Frith, 2003) and representation of mental states (Saxe, 2006).

Particularly intriguing in the aforementioned contrast was the activation in the hippocampi, which points to increased access to previous episodic experiences (Awad, Warren, Scott, Turkheimer, & Wise, 2007; Gilboa, Winocur, Grady, Hevenor, & Moscovitch, 2004; Lackner, 1974). Although the retrieval of past experiences is of key importance in mentalizing, hippocampal activation is not always seen to increase specifically for ToM tasks. Combined with the medial PFC, and connected via the retrosplenium (Vann, Aggleton, & Maguire, 2009), this points to an increase in the active comparison of stimuli with past memories in the determination of authenticity. Since this is contrasted with the emotion task, it would imply that this active comparison with self-/social experiences is of greater importance in the determination of social context than in the judgment of the emotion itself. However, although the contrast of task type included RT as a per-trial weighting and as a random effect, this cannot completely eliminate a potential confounding influence of task difficulty. Potential differences in task difficulty may induce a similar pattern in the so-called default network, which overlaps with part of the activation seen in the task-based contrast.

Although the task-based effects were of additional interest, the focus of the present study was the influence

of stimulus authenticity on emotion recognition and correlated brain activation. Since modulation by stimulus authenticity was examined for each task type separately, this analysis contrasts trials of identical tasks and similar task difficulty and eliminates the difficulty confounds mentioned previously. The ToM network was shown (based on a priori ROIs) to be modulated by differences in authenticity, both during emotion and authenticity judgments. Contrasting authentic versus play-acted stimuli showed up-regulation by authentic recordings in the medial PFC and retrosplenium, whereas discrimination of authenticity was shown to be quite poor through the signal detection analysis ($d' = 0.62$: difficulty recognizing authenticity), indicating that authenticity may have a greater influence than is immediately apparent simply from behavioral data. Although play-acted stimuli were hypothesized to cause greater activation because of the actor's intention to express the instructed emotion, the contrast of authentic versus play-acted emotions showed the opposite effect. In terms of modulation of mPFC activity, this may be related to its role in integrating various processing streams through its involvement in the representation of mental states (Gallagher & Frith, 2003), communicative intent (C. Frith & Frith, 2007), and self-referential thought (Ochsner & Gross, 2005). Participants' lack of knowledge of the exact context of the authentically spoken text may actually induce greater activation as opposed to play-acted recordings for which, if recognized, the intent is known. In this case, it is not a difference in intentionality of stimulus production that requires greater activation of the mPFC, but a lack of knowledge about the context in which that stimulus was originally produced.

Although differences in production context could have a significant impact during the authenticity task, this would not necessarily be the case during the emotion task, since recognition of the emotion expressed may not necessarily require knowledge of the intention behind that expression. Modulation of the retrosplenium points to a second, nonmutually exclusive explanation for the present activation patterns. Authenticity may lead to significantly different integration of (mPFC), and greater access to (retrosplenial cortex: Vann et al., 2009) previous experiences. This would suggest that individuals integrate recollections of their own experiences more for judgments involving authentic stimuli than for play-acted stimuli. The higher variability of previous authentic experiences thereby leads to greater activation when accessed, whereas play-acted emotions have been experienced less often and in less variable contexts as they are restricted to settings involving acting. Therefore, activation is up-regulated when the stimulus allows for more open-ended assumptions about social context, as with authentic emotions, whereas for play-acted recordings, acting represents the social context

and would thereby not induce an increase in ToM activation.

This effect could also explain how a bias toward identifying play-acted stimuli as angry, and authentic stimuli as sad (as uncovered through the choice theory analysis), could occur with relatively low recognition rates for both emotion (in comparison with other studies) and authenticity (as compared with chance rates). Although explicit discrimination of authenticity is quite difficult, though largely unbiased in this study (small criterion value from SDT), it nevertheless differentially modulates activity in the ToM network to bias the recognition of emotion. Therefore, authentic and play-acted stimuli have the potential to differentially influence experiments on both behavior and brain activation, even in tasks not explicitly related to stimulus authenticity. Although an acoustic analysis of all recordings (Jürgens, Hammerschmidt, & Fischer, 2011) comparing the authentic and play-acted sets showed that the two groups of recordings (original, reenacted) did not differ significantly in pitch (Hz: 188.77 ± 66.45 ; 188.7 ± 64.61), speech rate (syllables/s: 5.62 ± 1.62 ; 5.66 ± 1.66), or harmonic-to-noise ratio (mean: 0.525 ± 0.081 ; 0.517 ± 0.084), pitch variability and specific vowel voice quality parameters were found to correlate significantly with recording authenticity. Further research will be required to determine which of these stimulus properties are relevant to the authenticity effects discussed here.

Contrasting authentic versus play-acted stimuli also elucidated further authenticity-modulated activity in addition to the network described previously. Increased activation in anterior middle temporal cortex during the playback of authentic recordings similarly points to the wide range of possible social contexts available for authentically expressed emotions, accentuated by the higher level of integration implied by more anterior activation. Although the involvement of the anterior temporal cortex has been related specifically to prosody recognition (Adolphs, Damasio, & Tranel, 2002; Ethofer et al., 2006; Wiethoff et al., 2008), this specific link may be more generally due to its role in social cue perception in general, as was first demonstrated by lesions in the macaque monkey (Bachevalier & Meunier, 2005) and more specifically for social emotions in humans (Burnett & Blakemore, 2009).

Concluding remarks

Although play-acted stimuli could be expected to cause greater activation in the ToM network than authentic stimuli because of the speaker's additional intention to act, it instead appears that some element of the authentic recordings, as compared with play-acted, increased activation in parts of the

ToM network and in social-context processing. This effect may be driven by the greater social significance that authentic emotions have in real life (as compared with play-acted emotions) or, alternatively, authentic emotions may just be more potent than play-acted emotions in triggering spontaneous mentalizing. Combining these behavioral and functional results clarifies that the emotional authenticity of prosody is an important property influencing human responses to such stimuli. This has important implications for future research on human emotions and should be taken into account for studies using play-acted prosody.

Author Note M.D. would like to thank Franziska Korb for her exceptional Presentation and Perl scripting introductions and her continued help in ironing out programming kinks. We also thank Ralph Pirow for his course in “R” and guidance in the complex world of statistics.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Abraham, A., Werning, M., Rakoczy, H., von Cramon, D., & Schubotz, R. (2008). Minds, persons, and space: An fMRI investigation into the relational complexity of higher-order intentionality. *Consciousness and Cognition*, *17*, 438–450.
- Adolphs, R. (2009). The social brain: Neural basis of social knowledge. *Annual Review of Psychology*, *60*, 693–716.
- Adolphs, R., Damasio, H., & Tranel, D. (2002). Neural systems for recognition of emotional prosody: A 3-D lesion study. *Emotion*, *2*, 23–51.
- Agresti, A. (2007). *An introduction into categorical data analysis*. New Jersey: Wiley.
- Awad, M., Warren, J., Scott, S., Turkheimer, F., & Wise, R. (2007). A common system for the comprehension and production of narrative speech. *Journal of Neuroscience*, *27*, 11455.
- Bach, D., Grandjean, D., Sander, D., Herdener, M., Strik, W., & Seifritz, E. (2008). The effect of appraisal level on processing of emotional prosody in meaningless speech. *NeuroImage*, *42*, 919–927.
- Bachevalier, J., & Meunier, M. (2005). The neurobiology of social-emotional cognition in nonhuman primates. In A. Easton & N. Emery (Eds.), *The Cognitive Neuroscience of Social Behaviour* (1st ed., pp. 19–57). East Sussex, England: Psychology Press.
- Bänziger, T., & Scherer, K. (2007). Using actor portrayals to systematically study multimodal emotion expression: The GEMEP corpus. *Lecture Notes in Computer Science*, *4738*, 476–487.
- Belin, P., Fillion-Bilodeau, S., & Gosselin, F. (2008). The Montreal Affective Voices: A validated set of nonverbal affect bursts for research on auditory affective processing. *Behavior Research Methods*, *40*, 531–539.
- Bolton, T. (1902). A biological view of perception. *Psychological Review*, *9*, 537–548.

- Buchanan, T., Lutz, K., Mirzazade, S., Specht, K., Shah, N., Zilles, K., & Jäncke, L. (2000). Recognition of emotional prosody and verbal components of spoken language: An fMRI study. *Cognitive Brain Research*, *9*, 227–238.
- Buckner, R., Andrews-Hanna, J., & Schacter, D. (2008). The brain's default network: Anatomy, function, and relevance to disease. *Annals of the New York Academy of Sciences*, *1124*, 1–38.
- Burnett, S., & Blakemore, S. (2009). Functional connectivity during a social emotion task in adolescents and in adults. *European Journal of Neuroscience*, *29*, 1294–1301.
- Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., & Taylor, J. (2001). Emotion recognition in human–computer interaction. *IEEE Signal Processing Magazine*, *18*, 32–80.
- Ekman, P., Davidson, R., & Friesen, W. (1990). The Duchenne smile: Emotional expression and brain physiology II. *Journal of Personality and Social Psychology*, *58*, 342–353.
- Ekman, P., Friesen, W., & O'Sullivan, M. (1988). Smiles when lying. *Journal of Personality and Social Psychology*, *54*, 414–420.
- Ethofer, T., Anders, S., Erb, M., Herbert, C., Wiethoff, S., Kissler, J., et al. (2006). Cerebral pathways in processing of affective prosody: A dynamic causal modeling study. *NeuroImage*, *30*, 580–587.
- Everhart, D., Demaree, H., & Shipley, A. (2006). Perception of emotional prosody: Moving toward a model that incorporates sex-related differences. *Behavioral and Cognitive Neuroscience Reviews*, *5*, 92.
- Frick, R. (1985). Communicating emotion: The role of prosodic features. *Psychological Bulletin*, *97*, 412–429.
- Friston, K., Fletcher, P., Josephs, O., Holmes, A., Rugg, M., & Turner, R. (1998). Event-related fMRI: Characterizing differential responses. *NeuroImage*, *7*, 30–40.
- Frith, U., & Frith, C. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *358*, 459–473.
- Frith, C., & Frith, U. (2007). Social cognition in humans. *Current Biology*, *17*, 724–732.
- Gallagher, H., & Frith, C. (2003). Functional imaging of “theory of mind”. *Trends in Cognitive Sciences*, *7*, 77–83.
- Gilboa, A., Winocur, G., Grady, C., Hevenor, S., & Moscovitch, M. (2004). Remembering our past: Functional neuroanatomy of recollection of recent and very remote personal events. *Cerebral Cortex*, *14*, 1214.
- Goldstein, T. (2009). Psychological perspectives on acting. *Psychology of Aesthetics, Creativity, and the Arts*, *3*, 6–9.
- Grezes, J., Berthoz, S., & Passingham, R. (2006). Amygdala activation when one is the target of deceit: Did he lie to you or to someone else? *NeuroImage*, *30*, 601–608.
- Gur, R. C., Gunning-Dixon, F., Bilker, W., & Gur, R. E. (2002). Sex differences in temporo-limbic and frontal brain volumes of healthy adults. *Cerebral Cortex*, *12*, 998.
- Hall, J., & Schmid Mast, M. (2007). Sources of accuracy in the empathic accuracy paradigm. *Emotion*, *7*, 438–446.
- Hietanen, J., Surakka, V., & Linnankoski, I. (1998). Facial electromyographic responses to vocal affect expressions. *Psychophysiology*, *35*, 530–536.
- Hoekert, M., Vingerhoets, G., & Aleman, A. (2010). Results of a pilot study on the involvement of bilateral inferior frontal gyri in emotional prosody perception: An rTMS study. *BMC Neuroscience*, *11*, 93.
- Jürgens, R., Hammerschmidt, K., & Fischer, J. (2011). Authentic and play-acted vocal emotion expressions reveal acoustic differences. *Frontiers in Emotion Science*, *2*, 180.
- Kochunov, P., Mangin, J., Coyle, T., Lancaster, J., Thompson, P., Rivière, D., et al. (2005). Age-related morphology trends of cortical sulci. *Human Brain Mapping*, *26*, 210–220.
- Krueger, F., Barbey, A., & Grafman, J. (2009). The medial prefrontal cortex mediates social event knowledge. *Trends in Cognitive Sciences*, *13*, 103–109.
- Lackner, J. (1974). Observations on the speech processing capabilities of an amnesic patient: Several aspects of HM's language function. *Neuropsychologia*, *12*, 199.
- Laukka, P. (2005). Categorical perception of vocal emotion expressions. *Emotion*, *5*, 277–295.
- Lohmann, G., Müller, K., Bosch, V., Mentzel, H., Hessler, S., Chen, L., et al. (2001). Lipsia—a new software system for the evaluation of functional magnetic resonance images of the human brain. *Computerized Medical Imaging and Graphics*, *25*, 449–457.
- McClure, E. (2000). A meta-analytic review of sex differences in facial expression processing and their development in infants, children, and adolescents. *Psychological Bulletin*, *126*, 424–453.
- Norris, D. (2000). Reduced power multislice MDEFT imaging. *Journal of Magnetic Resonance Imaging*, *11*, 445–451.
- Ochsner, K., & Gross, J. (2005). The cognitive control of emotion. *Trends in Cognitive Sciences*, *9*, 242–249.
- Povinelli, D., & Preuss, T. (1995). Theory of mind: Evolutionary history of a cognitive specialization. *Trends in Neurosciences*, *18*, 418–424.
- Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *The Behavioral and Brain Sciences*, *1*, 515–526.
- R Development Core Team. (2008). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org>
- Rakoczy, H., & Tomasello, M. (2006). Two-year-olds grasp the intentional structure of pretense acts. *Developmental Science*, *9*, 557–564.
- Saxe, R. (2006). Uniquely human social cognition. *Current Opinion in Neurobiology*, *16*, 235–239.
- Scheiner, E., & Fischer, J. (2011). Emotion expression: The evolutionary heritage in the human voice. In W. Welsch (Ed.), *Interdisciplinary anthropology: The continuing evolution of man*. Heidelberg, Germany: Springer.
- Scherer, K. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin*, *99*, 143–165.
- Schirmer, A., & Kotz, S. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences*, *10*, 24–30.
- Schirmer, A., Kotz, S., & Friederici, A. (2005). On the role of attention for the processing of emotions in speech: Sex differences revisited. *Cognitive Brain Research*, *24*, 442–452.
- Smith, J. E. K. (1982). Recognition Models Evaluated: A Commentary on Keren and Baggen. *Perception & Psychophysics*, *31*, 183–189.
- Spreng, R., Mar, R., & Kim, A. (2009). The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: A quantitative meta-analysis. *Journal of Cognitive Neuroscience*, *21*, 489–510.
- Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain: 3-dimensional proportional system: An approach to cerebral imaging*. New York, NY: Thieme.
- Ugurbil, K., Garwood, M., Ellermann, J., Hendrich, K., Hinke, R., Hu, X., et al. (1993). Imaging at high magnetic fields: Initial experiences at 4T. *Magnetic Resonance Quarterly*, *9*, 259–277.
- Van Overwalle, F. (2009). Social cognition and the brain: A meta-analysis. *Human Brain Mapping*, *30*, 829–858.
- Van Overwalle, F., & Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: A meta-analysis. *NeuroImage*, *48*, 564–584.
- Vann, S. D., Aggleton, J. P., & Maguire, E. A. (2009). What does the retrosplenial cortex do? *Nature Reviews Neuroscience*, *10*, 792–802.
- Wiethoff, S., Wildgruber, D., Kreifelts, B., Becker, H., Herbert, C., Grodd, W., et al. (2008). Cerebral processing of emotional

- prosody—influence of acoustic parameters and arousal. *NeuroImage*, 39, 885–893.
- Wildgruber, D., Hertrich, I., Riecker, A., Erb, M., Anders, S., Grodd, W., et al. (2004). Distinct frontal regions subserve evaluation of linguistic and emotional aspects of speech intonation. *Cerebral Cortex*, 14, 1384.
- Wildgruber, D., Riecker, A., Hertrich, I., Erb, M., Grodd, W., Ethofer, T., et al. (2005). Identification of emotional intonation evaluated by fMRI. *NeuroImage*, 24, 1233–1241.
- Wilting, J., Kraemer, E., & Swerts, M. (2006). Real vs. acted emotional speech. In *INTERSPEECH 2006 and 9th International Conference on Spoken Language Processing, INTERSPEECH 2006—ICSLP Pittsburgh, PA* (Vol. 2, pp. 805-808).
- Worsley, K., & Friston, K. (1995). Analysis of fMRI time-series revisited—again. *NeuroImage*, 2, 173–181.
- Wurm, M., von Cramon, D., & Schubotz, R. (2011). Do we mind other minds when we mind other minds' actions? A functional magnetic resonance imaging study. *Human Brain Mapping*.