Explicit authenticity and stimulus features interact to modulate BOLD response induced by emotional speech

Matthis Drolet · Ricarda I. Schubotz · Julia Fischer

Abstract Context has been found to have a profound effect on the recognition of social stimuli and correlated brain activation. The present study was designed to determine whether knowledge about emotional authenticity influences emotion recognition expressed through speech intonation. Participants classified emotionally expressive speech in an fMRI experimental design as sad, happy, angry, or fearful. For some trials, stimuli were cued as either authentic or play-acted in order to manipulate participant top-down belief about authenticity, and these labels were presented both congruently and incongruently to the emotional authenticity of the stimulus. Contrasting authentic versus play-acted stimuli during uncued trials indicated that play-acted stimuli spontaneously up-regulate activity in the auditory cortex and regions associated with emotional speech processing. In addition, a clear interaction effect of cue and stimulus authenticity showed up-regulation in the posterior superior temporal sulcus and the anterior cingulate cortex, indicating that cueing had an impact on the perception of authenticity. In particular, when a cue indicating an authentic stimulus was followed by a play-acted stimulus, additional activation occurred in the temporoparietal junction, probably pointing to increased load on perspective taking in such trials. While actual authenticity has a significant impact on brain activation, individual belief about stimulus authenticity can additionally modulate the brain response to differences in emotionally expressive speech.

Keywords Emotion recognition · Speech · Prosody · Theory of mind · Context · Authenticity

Introduction

Perception of social stimuli, such as facial expressions, gestures, and vocal expressions of emotions, is a key component in human daily life. These stimuli may vary strongly in their content and ambiguity and frequently require the integration of multiple stimulus modalities (Ethofer, 2006; Hornak, Rolls, & Wade, 1996; Kreifelts, Ethofer, Grodd, Erb, & Wildgruber, 2007; Regenbogen et al., 2012; Scherer & Ellgring, 2007). Much research in this area focuses on the variation in stimulus quality and bottom-up influence on perception, while not addressing top-down modulation (Banse & Scherer, 1996; Pell, Monetta, Paulmann, & Kotz, 2009; Scherer, 1991). Context may, however, have a profound top-down influence on the recognition of social stimuli (Nelson & Russell, 2011; Pell, 2005). Albert, Lópex-Martin, and Carretié (2010) showed that stimuli portrayed with positive emotional context induced greater activation during response inhibition in an ERP experimental design, while Schirmer (2010) showed that valence of neutral words in memory is influenced by the affective prosody present during encoding but does not affect retrieval ability. Much of this research is focused on the bottom-up effect of perceiver emotional state (Blanchette & Richards, 2010) or stimulus emotional content (Barrett & Kensinger, 2010; Chung & Barch, 2011), while, conversely, the influence of

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

M. Drolet · R. I. Schubotz
Biological Psychology, University of Münster, Münster, Germany

M. Drolet · R. I. Schubotz
Minerva Group Motor Cognition, Max Planck Institute
for Neurological Research, Cologne, Germany

higher cognition on emotion perception has yet to be explored in detail.

To examine the overall integration of both top-down and bottom-up processes, Teufel and colleagues proposed the theory of social perception as a conceptual framework (Teufel, Fletcher, & Davis, 2010). Central to this framework is theory of mind (ToM; Premack & Woodruff, 1978), with its role in interpreting social stimuli by ascribing thoughts and intentions to other individuals (Hooker, Verosky, Germine, Knight, & D'Esposito, 2008; Mier et al., 2010; O'Brien et al., 2011). The brain network responsible for these processes includes the medial prefrontal cortex (mPFC), posterior superior temporal sulcus (pSTS), temporoparietal junction (TPJ), and retrosplenial cortex (U. Frith & Frith, 2003). The importance of the integration of emotion and ToM can be seen in research on empathy (Shamay-Tsoory, 2011; Vollm et al., 2006) and affective ToM (Mier et al., 2010). However, here too the influence of ToM on the perception of emotional stimuli remains largely unexamined.

Drolet, Schubotz, and Fischer (2011) showed that authenticity can affect the perception of the emotion expressed. We contrasted speech produced without external instruction (*authentic*) with that produced by professional actors after instruction (*play-acted*). The ability to feign emotions and its relation to ToM is known from play behavior (Rakoczy, Tomasello, & Striano, 2004; Schmidt, Pempek, Kirkorian, Lund, & Anderson, 2008) and deceit (DePaulo, 2003; Grezes, Berthoz, & Passingham, 2006; Grezes, Frith, & Passingham, 2004). Since the detection of play-acting involves the perception of the speaker's intention, the perception of acted emotions is an ideal case of the interaction of emotion perception and higher cognition. In addition, while professional acting is of major social and economic impact in entertainment and research, its impact on perception has not been thoroughly examined (Goldstein, 2009, 2011).

The present study was designed to examine the effects of assumed authenticity on emotion perception by labeling stimuli with contextual information. The aim was to determine whether assumptions about the authenticity of emotional speech intonation (prosody) bias the recognition of the expressed emotion. The effect of emotional prosody on behavioral responses is well known (Buchanan et al., 2000; Jaywant & Pell, 2011; Scherer, 1991), and brain activation correlated with its perception is an active area of research (Kotz, Kalberlah, Bahlmann, Friederici, & Haynes, 2012; Vigneau, 2006; Wildgruber, Pihan, Ackermann, Erb, & Grodd, 2002) Drolet et al. (2011) showed that anger was recognized best when play-acted, while sadness was recognized best when authentic, and, using choice theory (Kornbrot, 1978; Luce, 1959), we found a bias, such that authentic stimuli were more often categorized as sad and play-acted ones were more often categorized as angry, overall. BOLD response analyses showed that both active

discrimination of authenticity and authentic stimuli (vs. play-acted) up-regulated activation in the ToM network (medial prefrontal, retrosplenial, and temporoparietal cortices). Accordingly, we suggested that ToM network activity is induced in the listener by differences in authenticity of vocal emotional expression. What remained unclear, however, was whether this effect was purely stimulus driven (bottom-up) or additionally modulated top-down by assumptions about stimulus authenticity. In the former case, emotional intonation itself may change the listener's perception of stimulus emotion directly. In the latter, acoustic differences may signal authenticity, the perception of which engages ToM as top-down modulation of perception in the listener, which then influences emotion categorization. Therefore, the study presented here was designed to determine whether the hypothesis of the top-down effects of stimulus authenticity could be corroborated.

In order to do so, we used a modified paradigm in which participants only categorized stimuli by emotion. To manipulate top-down modulation, we labeled stimuli on-screen as authentic or play-acted prior to playback, both congruently and incongruently to the stimulus. If the effect of authenticity on emotion categorization is purely a bottom-up effect of differences in emotional expression, these labels should have no effect on recognition rates. If, however, this effect is induced top-down, labels should alter or even override the stimulus-induced authenticity effect. More specifically, authentic cues should decrease recognition of anger and increase recognition of sadness (Drolet et al., 2011), irrespective of actual stimulus authenticity. Second, due to the involvement of ToM as previously shown, we expected authentic cues to up-regulate activity in the mPFC, the retrosplenial cortex and the TPJ (U. Frith & Frith, 2003) if the effects were induced top-down. However, if authenticity affects emotion recognition bottom-up, we expected authentic stimuli to up-regulate BOLD, as was found by Drolet and colleagues, as well as regions associated with emotional prosody perception, including the orbital and inferior frontal, superior temporal, and inferior parietal cortices (as reviewed by Schirmer & Kotz, 2006). Activation in the cingulate cortex was also predicted due to its involvement in both emotion perception and ToM processes (Bach et al., 2008; Buchanan et al., 2000; Wildgruber et al., 2005).

Method

Participants

Eighteen female participants (20–30 years of age, $M = 24$ years; right-handed; mother tongue, German) were contacted through the Cologne MPI database for fMRI experiments. Only individuals without neurological or psychological complications (including the use of psychiatric medication)

were selected. 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 remunerated. The Ethics Commission of Cologne University's Faculty of Medicine approved the experimental standards and procedure, and data were handled pseudony-mously, with identity blind to the experimenter.

Stimulus selection

Original recordings (mono wave format; sample rate of 44.1 kHz) were selected from German radio archives. These were interviews with individuals talking in an emotional fashion (anger, fear, joy, or sadness) about a highly charged ongoing or recollected event (e.g., the death of a child, winning a lottery, in rage about injustice, or threatened by a current danger). Emotion was ascertained through verbal content and context described in recording summaries. Staged settings (e.g., talk shows) and scripted interviews were excluded. Of the 80 speech tokens 35 were made outdoors and varied in their noise surroundings, but were of good quality with minimal background noise. To ensure inference-free verbal content, naïve participants rated text-only transcripts. Recordings with emotion recognized better than chance were replaced to ensure that semantic content would not indicate emotional content. The original set con-sisted of 20 samples per emotion (half male, half female; total of 80 recordings by 78 speakers; mean, 1.75 s ± 1.00 SD; range, 0.36–4.06 s). The average amplitude of all stim-uli was equalized with Avisoft SASLab Pro Recorder Ver-sion 4.40 (Avisoft Bioacoustics, Berlin, Germany).

Actors from Germany performed the play-acted stimuli (42 actors each replicated a maximum of three recordings of equivalent emotional content). The actors were told to ex-press 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). Play-acted re-cording environment was varied by location while avoiding excessive background noise (mean, 1.76 s ± 1.02 SD; range, 0.38–4.84 s). The average amplitude of all stimuli was equalized with Avisoft SASLab Pro Recorder Version 4.40 (Avisoft Bioacoustics, Berlin, Germany).

Trial and stimulus presentation

NBS Presentation (Neurobehavioral Systems, Inc., Albany, California) controlled trial sequence and the timing of each experimental run. Each participant completed $n = 178$ trials (limited by the number of available stimuli with no stimulus repetition). $N = 144$ stimuli were used for the emotion judgment task. One third ($n = 48$) were not cued (no context information was provided), one third were cued as

authentic, and one third were cued as play-acted. Of the cued trials, half of the authentic and play-acted cues were congruent to the stimulus presented in the trial, while the others were incongruent. For emotion judgments, four responses were possible: anger, sadness, happiness, and fear (presented in German as "Wut," "Trauer," "Freude," and "Angst"). Cues were presented above the response options as authentic or played ("echt" or "spiel," respectively). The remaining 30 trials were used to implement two independent control tasks: 18 empty trials with pink-noise playback, and 16 age task trials (with eight authentic and eight play-acted stimuli not used during experimental trials) in which partic-ipants had to determine the age of the speaker. Speaker age determination was selected as a behavioral control task due to its similarity to the emotion judgment task in relying on prosodic information (Linville, 1985; Schötz, 2007). Partic-ipants were tasked to select the age closest to the age of the speaker (20, 30, 40, or 50).

Trial type and stimulus type were pseudorandomized to reduce any systematic effects that could have otherwise occurred with simple randomization. Each participant was shown a button sequence on-screen (800 × 600 pixel video goggles; NordicNeuroLab, Bergen, Norway) complementa-ry to the response box layout (10 × 15 × 5 cm gray plastic box with a row of four black plastic buttons). All buttons were assigned a possible response for the emotion judgment and age categorization tasks. The maximal line-of-sight angle for visual information was kept under 5° to minimize eye movement (Fig. 1).

Experimental procedure

Participants were fitted with headphones for audio playback (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 (Co-logne, Germany) system equipped with a standard birdcage head coil. Participants were placed with their four fingers (excluding thumb) positioned on the response buttons of the response box. Form-fitting cushions were utilized to prevent head, arm, and hand movements. Twenty-four axial slices (210-mm field of view; 64 × 64 pixel matrix; 4-mm thick-ness; 1-mm spacing; in-plane resolution of 3 × 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, 24 anatom-ical 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
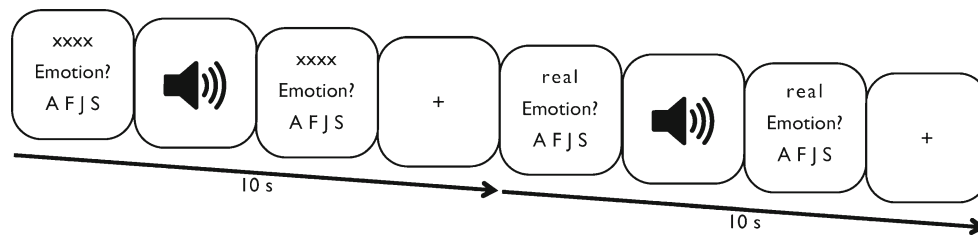
Fig. 1 Experimental trial sequences for uncued and cued emotion judgment trials with durations (in seconds). A F J S represent the four possible emotion responses (forced choice design: anger, fear, joy, and sadness, respectively). xxxx indicates display content during uncued trials

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 (duration of 1 min), one experimental run lasted approximately 45 min.

Behavioral statistical analysis

Recognition accuracy was analyzed using the R Statistical Package v2.15 (R Development Core Team, 2008). The generalized linear model was implemented to determine the best model fit for recognition rates using the glmer function from the lme4 package with binomial error structure and logit link function (GLMM; Baayen, Davidson, & Bates, 2008). The basic model examined the effect of the independent variables (stimulus emotion, stimulus authenticity, and cue type) on the response variable (whether emotion rating was correct or not), with participant and speaker repetition included as random effects [glmer(Rating~Emotion * Authenticity * Cue + (1|Participant) + (1|Repetition))]. First, the overall influence of each categorical independent variable was examined via a likelihood ratio test with a chi-squared distribution (χ^2). Models with and without the effect of interest were compared using the ANOVA function, indicating the overall variable effect (main effects and interactions). If a significant effect was found, models were examined for interactions (main effects were ignored). If no interaction was found, main effects of the respective independent variables on the response variable were examined. The effects of each categorical independent variable or, if appropriate, interactions between variables are indicated by χ^2 values, degrees of freedom, and effect significance (*p*-value). Mean behavioral data are presented as reaction times (RTs, in seconds), as well as the respective probability of correct responding for each emotion, authenticity, and cue category, with 95 % confidence intervals.

Functional MRI statistical analysis

After motion correction using Siemens rigid-body registration protocol (München, Germany), 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 (www.cbs.mpg.de/institute/). 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 set for each scanned participant dependent on the pseudorandomized design (filter frequency range: 1/75–85 Hz). Spatial smoothing was performed with a Gaussian filter of 5.65 mm FWHM (sigma = 2.4). 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 applied. 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 data set 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 × 3 × 3 mm (27 mm^3).

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, with trials classified by all levels of the three main factors emotion category, cueing type, and stimulus authenticity was generated with a delta function, convolved with the hemodynamic response function (gamma function). Each trial in the design matrix was identified by onset time and stimulus length, while speaker repetition was included as REG No INT to prevent this from influencing the statistical analysis. 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 (*M* = 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 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-participant contrast images were entered into a second-level random effects analysis for each of the contrasts. One-sample t-tests (two-sided) 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 correct for false-positive results, first an initial voxel-wise z-threshold was set to $z = 2.58$ ($p < .01$, two-sided). In a second step, the results were corrected for multiple comparisons at the cluster level using cluster-size and cluster-value thresholds obtained by Monte Carlo simulations at a significance level of $p < .05$ (i.e., the reported activations are significantly activated at $p < .05$).

BOLD contrasts were computed on the basis of task and stimulus categories. Trials with authentic stimulus playback were contrasted against trials with play-acted stimulus playback. First, a contrast within only noncued trials examined the hypothesis-driven main effect of authenticity, independently of cueing. Second, the main effect of cueing was examined by contrasting cued versus uncued trials. Then, to determine how cueing modulates authenticity-induced activation, the interaction effect of cueing and authenticity was examined utilizing a 2 × 2 factorial design. Trials were labeled according to both stimulus authenticity and cue type, producing a set of four experimental trial categories to be analyzed: (1) authentic stimulus, authentic cue; (2) play-acted stimulus, authentic cue; (3) authentic stimulus, play-acted cue; and (4) play-acted stimulus, play-acted cue. In this design, the contrast to indicate activation due to interaction of authenticity and cue is [authentic stimulus, authentic-cue > play-acted stimulus, authentic cue] > [authentic stimulus, play-acted cue > play-acted-stimulus, play-acted cue]. This contrast shows the effect cue validity has on the effect of stimulus authenticity. This simultaneously represents a contrast of incongruent and congruent trials, so that the interaction effects of stimulus and cue can be understood in terms of the effects of cue congruency on stimulus perception. In order to determine the direction of the interaction effects found in this contrast, authentic stimuli were contrasted against play-acted stimuli for trials with authentic and play-acted cues individually in the form of post hoc t-tests. Due to the complexity of the effects of interest in this hypothesis-driven analysis, small-volume correction was performed using a restricted search volume (spheres with a 20-voxel radius), as well as cluster-size and cluster-value thresholds obtained by Monte Carlo simulations. 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) based on the review by Van Overwalle and Baetens (2009); (Wurm, von Cramon, & Schubotz, 2011) for ToM and the review by Vogt (2005) for the cingulate cortex, were summed to calculate the alpha level of each individual activation cluster within the a priori hypothesized regions of the ToM network. Therefore, analyses simultaneously included all voxels from the coordinates from all hypotheses.
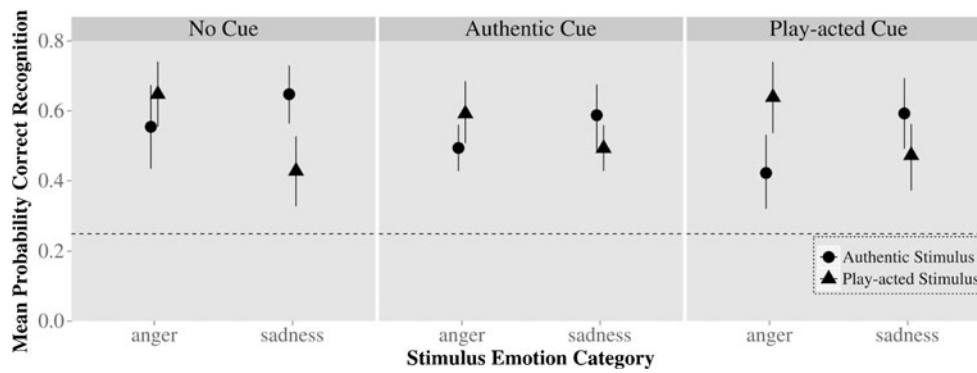
Results

Behavioral results

First, the effect of cueing was examined as an independent variable in the model (three levels: no cue, authentic cue, play-acted cue). This model, including all four emotion categories (anger, fear, joy, sadness), stimulus authenticity (authentic, play-acted), and cueing, indicated that cueing had no significant effect on emotion categorization, $\chi^2(15) = 16$, $p > .10$. Since the effect of authenticity on emotion recognition was predicted only for anger and sadness stimuli, we examined a reduced model that contained only those two emotions to ensure that any potentially weak effects would not be overlooked. Overall, however, this model also indicated that cueing had no influence on recognition in interaction with either authenticity and emotion [model including three-way interaction: $\chi^2(6) = 8.43$, $p > .10$] or emotion only [cueing and emotion: $\chi^2(6) = 5.91$, $p > .10$]. While effect sizes are different between cueing categories, none of these differences are significant and are unlikely to be systematic (Fig. 2). Categorization rates and RTs are presented in Table 1, split by trial type.

To ensure that the present results are otherwise consistent with previous findings, effects of stimulus authenticity on emotion categorization only (not split by cueing category) were analyzed using the generalized linear mixed model. Overall, participants detected anger correctly more often when play-acted than when authentic, whereas they detected sadness correctly more often when authentic than when play-acted [significant emotion × authenticity interaction, $\chi^2(3) = 28.63$, $p < .001$; see Fig. 3]. Participants detected fear only near chance levels for both authentic and play-acted recordings. Post hoc analyses indicated that recognition of joy was significantly better than recognition of fear ($p < .05$) but was not affected by authenticity ($p > .10$).

We also modeled RTs statistically to determine whether emotion, authenticity, or cueing influences behavior other than emotion categorization. Analyses indicated a significant main effect of stimulus emotion, $\chi^2(3) = 76.11, p < .01$, but none of stimulus authenticity, $\chi^2(1) = 0.08, p > .10$, or cueing, $\chi^2(2) = 2.02, p > .10$, on RTs. Considering the

Fig. 2 Behavioral results from fMRI for judgments on anger and sadness stimuli, separated by cueing type: Actual emotion of recordings on *x*-axis, probability of recognition of emotion on *y*-axis. Chance level of recognition is indicated by the dotted line. Circles indicate recognition rates for authentic stimuli, while triangles indicate recognition rates for play-acted stimuli, along with 95 % confidence intervals

average RTs that were measured from participant responses, it is unlikely that any differences in BOLD effects reported in the following were due to variability in participant concentration level or stimulus difficulty.

fMRI results

First, stimulus authenticity was contrasted only within uncued trials to determine the main effect of spontaneous modulation by authenticity (whole-brain analysis). Activation loci with significantly greater activation during trials without cueing [play-acted stimuli > authentic stimuli] included the transverse temporal gyrus (TTG), pSTS, TPJ, fusiform gyrus, inferior frontal gyrus, superior frontal gyrus, and middle frontal gyrus (Fig. 4, Table 2). The direct contrast cued versus uncued trials indicated no significant differences in activation between these two conditions.

The third analysis step examined the interaction effect between stimulus authenticity and cueing (small-volume analysis). As has been stated, this contrast applied a 2 × 2 design with stimulus authenticity versus cue validity to determine the effect of cueing interacting with stimulus authenticity over all trials. The stimulus × cue interaction contrast revealed activation in the pSTS, TPJ, and anterior cingulate cortex (ACC; Fig. 5; Table 3).

Finally, to examine these interaction effects in more detail, authentic stimuli were contrasted against play-acted stimuli post hoc, for trials with authentic and play-acted cues individually. For trials with authentic cues, this revealed significantly increased activation for play-acted stimuli in the TPJ and pSTS. Conversely, the same contrast performed for trials with only play-acted cueing did not show any significant activation (Fig. 5; Table 3). This empty *z*-map pointed to a relatively low power of the ACC effect revealed in the cue × stimulus authenticity interaction. When inspecting this *t*-test with lowered statistical threshold, we found the ACC to be (descriptively) more engaged for authentic, as compared with play-acted, stimuli when cueing was "play acted."

Table 1 Emotion categorization accuracy and reaction time for cue type

Recording	Type	No cue		Authentic cue		Play-acted cue	
		Recognition	RT	Recognition	RT	Recognition	RT
Authentic	Anger	0.56 ± 0.26	2.92 ± 0.55	0.49 ± 0.15	2.92 ± 0.51	0.42 ± 0.23	2.96 ± 0.60
	Fear	0.14 ± 0.14	3.06 ± 0.72	0.20 ± 0.13	3.14 ± 0.69	0.27 ± 0.15	3.17 ± 0.48
	Joy	0.35 ± 0.22	3.07 ± 0.50	0.40 ± 0.21	3.19 ± 0.68	0.30 ± 0.18	3.20 ± 0.48
	Sadness	0.65 ± 0.19	3.16 ± 0.62	0.59 ± 0.21	3.17 ± 0.75	0.59 ± 0.23	3.22 ± 0.73
Play-acted	Anger	0.65 ± 0.21	3.04 ± 0.72	0.59 ± 0.20	2.56 ± 0.57	0.64 ± 0.22	2.54 ± 0.34
	Fear	0.20 ± 0.21	2.48 ± 0.57	0.19 ± 0.13	3.23 ± 0.60	0.18 ± 0.13	3.35 ± 0.48
	Joy	0.32 ± 0.18	3.39 ± 0.51	0.32 ± 0.17	3.11 ± 0.61	0.33 ± 0.17	3.33 ± 0.54
	Sadness	0.43 ± 0.23	3.03 ± 0.65	0.49 ± 0.15	3.25 ± 0.73	0.47 ± 0.22	3.35 ± 0.56

Note. Recognition rates for emotions (Recognition) and reaction times (RTs) for the respective categories. Rows are split by actual stimulus authenticity and columns segregated by cue type. Emotion trials required a determination of the emotional content of the recordings
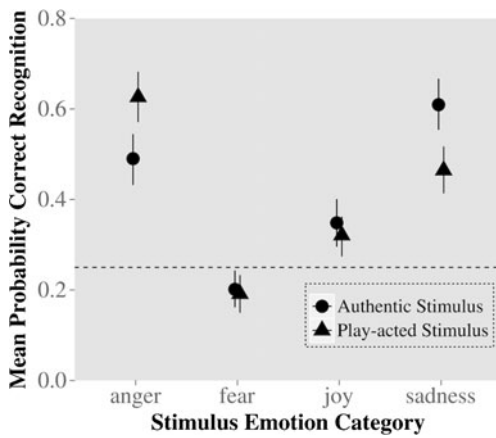
Fig. 3 Behavioral results from fMRI for emotion judgments: Actual emotion of recordings on *x*-axis, probability of recognition of emotion on *y*-axis. Chance level of recognition is indicated by the dotted line. Circles indicate recognition rates for authentic stimuli, while triangles indicate recognition rates for play-acted stimuli, along with 95 % confidence intervals

Table 2 Direct effects of stimulus authenticity (play-acted stimuli > authentic stimuli)

Area	BA	Hemisphere	Talairach coordinates			
			x	*y*	*z*	*Z*
No cue						
TTG	41	L	−38	−31	6	4.06
IFG	45	R	49	32	0	3.42
Anterior insula		L	−32	26	6	3.74
pSTS	21	R	49	−34	6	4.10
	22	L	−50	−28	0	3.63
Fusiform gyrus	37	R	43	−49	−12	3.63
mSFG	8	R	4	17	51	3.85
	6	L	−5	−1	54	3.51
MFG	6	L	43	2	39	3.57

Note. Anatomical specification, Talairach coordinates, and maximum *Z* value of local maxima ($p < .05$, corrected). TTG, transverse temporal gyrus; IFG, inferior frontal gyrus; pSTS, posterior superior temporal sulcus; mSFG, medial superior frontal gyrus; MFG, middle frontal gyrus.

Discussion

In this fMRI study, we investigated whether information about authenticity of emotional speech influenced perception of the respective emotions in a top-down manner. We examined stimuli varying in their emotional content and authenticity for their bottom-up effects on behavior and brain activation and assessed whether expectation (by cueing specific stimuli as either authentic or play-acted) had a top-down effect on emotion recognition and BOLD response. The behavioral data showed that the interaction effect of stimulus authenticity and emotion recognition is equivalent to that in Drolet et al. (2011). While cueing stimuli as either authentic or play-acted did not influence emotion recognition itself, the results of the BOLD response analysis included both expected and surprising effects of

cueing. The lack of effects in the behavioral data makes a case for a purely bottom-up effect of authenticity on emotion recognition. While the bottom-up effects are clearly of more importance in these experimental tasks, the examination of brain activation correlates shows that additional information is not completely ignored, as is discussed below.

A significant influence of authenticity on brain activity was observed by contrasting authentic versus play-acted stimuli for trials without context cues (Fig. 4). This contrast yielded an increase in activity in the TTGs and pSTS, extending into the TPJ on the right for play-acted, as opposed to authentic, stimuli. The lack of modulation in the complete ToM network in this contrast was surprising. This
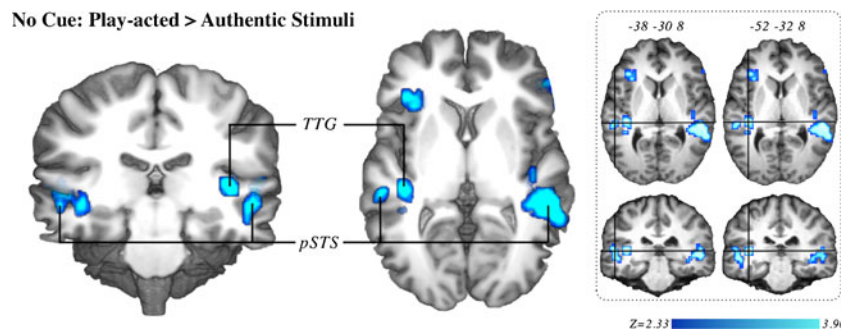
Fig. 4 Brain activation correlates of experimental tasks: Group-averaged (*n* = 18) statistical maps of significantly activated areas for play-acted versus authentic stimuli; corrected for false-positive results with an initial voxel-wise *Z*-threshold set to *Z* = 2.58 ($p < .01$, two-sided), and cluster-size and cluster-value thresholds obtained by Monte Carlo simulations at a significance level of $p < .05$. Activation was mapped onto the best average participant 3-D anatomical map. Left: anterior view coronal section through transverse temporal gyrus (TTG). Middle: top view sagittal section through TTG. Right: Two sections through TTG with crosshairs (centered on labeled coordinates) placed within pSTS (left) and TTG (right) in the left hemisphere. pSTS, posterior superior temporal sulcus

Fig. 5 Brain activation correlates of experimental tasks: Group-averaged ($n = 18$) statistical maps of significantly activated areas. **a** Interaction effects of cue and stimulus authenticity Signal change figures descriptively indicate the nature of the interaction based on the BOLD time-course from labeled regions in the z-map. **b** Contrasts of play-acted versus authentic stimuli during trials cued as authentic (top) or play-acted (bottom) to determine significance of interaction sources
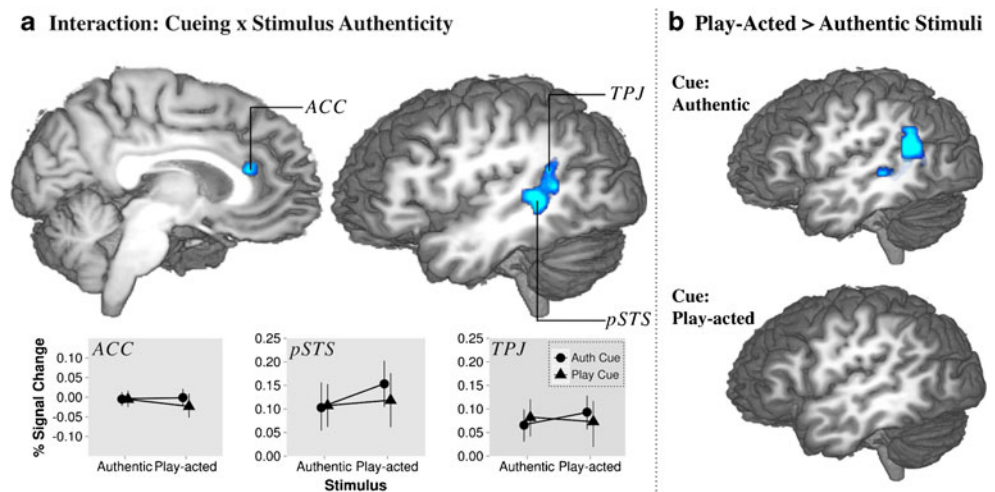
difference from Drolet et al. (2011) indicates that the effects seen in that study were partly task related. Therefore, the effect of stimulus authenticity depends on individual expectation and focus, as modulated by task. This is discussed in more detail below. While there are hemispheric differences in the present results, these are difficult to interpret. First, hemispheric differences are an active area of research, and several non-hypothesis-driven reverse inferences could be valid in this context. Second, corrections for multiple comparisons required for fMRI BOLD analyses can exaggerate the influence of laterality differences, while such apparent differences represent nonstatistical comparisons between p-values (differences between p-values may not be significant). Therefore, here these effects will not be discussed in detail.

The modulation in the TTG is indicative of an influence on early cortical auditory processing (Arnott, Binns, Grady, & Alain, 2004), likely due to differences in acoustic features between authentic and play-acted stimuli (Jürgens, Hammerschmidt, & Fischer, 2011). Therefore, the

Table 3 Effects of stimulus authenticity during cued trials

Area	BA	Hemisphere	Talairach coordinates			
			x	y	z	Z
Cueing × authenticity (interaction)						
pSTS	22	L	−47	−39	6	2.63
TPJ	39	L	−47	−52	12	3.08
ACC	24	R	7	26	18	3.48
Authentic cue: play-acted stimuli > authentic stimuli						
pSTS	22	L	−47	−43	3	2.55
TPJ	39	L	−50	−55	23	3.08

Note. Anatomical specification, Talairach coordinates, and maximum Z values of local maxima ($p < .05$, corrected). TPJ, temporoparietal junction; pSTS, posterior superior temporal sulcus; ACC, anterior cingulate cortex.

direct influence of authenticity on activation appears to be a spontaneous response to contextual differences between stimuli, particularly when participants are not told about authenticity or do not have to explicitly determine authenticity. Since this effect is stimulus based, it must be due to some acoustical difference between the categories of stimuli. Jürgens et al. found that a specific acoustic property, the contour or variability of the fundamental frequency ($f0$), is higher in play-acted than in authentic stimuli. A future study would be needed to determine whether $f0$ contour indeed modulates activity in the TTG in this case.

The TPJ is heavily involved in the representation of mental states and beliefs (Samson, Apperly, Chiavarino, & Humphreys, 2004; Saxe, 2006). As such, both these regions are essential to the network activated for ToM tasks. Drolet et al. (2011) showed that the TPJ is modulated by authenticity when participants categorize by authenticity explicitly, which requires perception of intention. Therefore, it is likely that the effect seen in this study is also related to representation of mental states (Samson et al., 2004). While Beaucousin et al. (2007) found that the left TPJ was not preferentially activated for emotional content and other studies have found the right TPJ to be more prominent in the perception of emotional prosody (Baum & Pell, 1999), the left TPJ does appear to be relevant when a combination of mental state perception and emotion recognition is required, including potential context-relevant causes for those states, as was hypothesized by Hervé, Razafimandimby, Vigneau, Mazoyer, and Tzourio-Mazoyer (2012).

Several studies have shown involvement of the pSTS for mentalizing about intentions in social situations (C. Frith, 2008) and representing human actions and relevant contexts (Saxe, 2006; Truett, Puce, & McCarthy, 2000). In particular, when contrasting tasks of cognitive versus affective ToM, Sebastian et al. (2012) found that only the pSTS, the temporal poles, and the precuneus were consistently activated for both, indicating a specific network overlap in these

regions for ToM and affective perception. However, another potential influence must be considered—in particular, due to the simultaneous modulation in the TTG. While primary auditory processing occurs in the TTG, further perception of lower-frequency activity, such as phonology or intonation, has been shown to correlate with activity in the lateral temporal gyrus (Beaucousin et al., 2007; Wildgruber et al., 2005). Here, the activation of the pSTS may be related to the perception of lower-frequency changes. As was mentioned above, Jürgens et al. (2011) found that $f0$ contour, a low-frequency variation in intonation, shows the strongest correlation with authenticity. While these two approaches to explaining pSTS activation may indicate some common process at work in both (Redcay, 2008), in this study, they are not mutually exclusive. Differences in intonation may lead to initial differential activation bottom-up, activating the pSTS even further when differences in intention or cause of action are perceived.

In order to examine top-down effects, trials were cued to examine whether the influence of explicitly stated authenticity and, hence, expectation would have a similar effect to those of actual authenticity as discussed above. Cues were presented both congruently and incongruently. Cues appear to have had little effect on the categorization results, while they were nevertheless incorporated into stimulus perception and influenced BOLD response. To determine whether cueing modulated brain activity during emotion perception, we examined the effect of the interaction of cue and stimulus authenticity. This contrast indicated that incongruent trials (authentic emotion following the announcement of play-acted emotion and vice versa) induced increased activation in the pSTS and TPJ (Fig. 5a). This interaction effect was found to be due to authentically cued play-acted emotions, as opposed to authentically cued authentic emotions, as shown in Fig. 5b. In addition, the interaction contrast indicated increased activation in the ACC (Fig. 5a). Activation in the ACC was also seen when contrasting authentic versus play-acted stimuli cued as play-acted, but this effect did not survive statistical correction and, hence, cannot be unambiguously interpreted at this point. However, on the basis of its presence in the full interaction contrast, it can be suggested that the ACC is slightly sensitive for incongruently cued emotional stimuli. This parallels studies that show that activation in this region correlates with stimulus conflict (MacDonald, Cohen, Stenger, & Carter, 2000) and uncertainty (Kéri, Decety, Roland, & Gulyás, 2004).

The interaction and post hoc contrasts (Fig. 5) show that increased activation in the TPJ and pSTS is due to a very specific set of circumstances. A prior belief of an authentic expression, with subsequent perception of a fake or play-acted one, recruits the area comprising the TPJ and pSTS. As was mentioned above, this stands in contrast to the ACC effect, which did not reach significance in the t-test. This

difference may indicate that an invalidly cued play-acted stimulus is much more salient than an invalidly cued authentic one (Fig. 5b). This salience may be due to the greater day-to-day relevance of believed authentic expressions shown not to be, as in the case of deception detection (Burgoon, Blair, & Strom, 2008), which is known to recruit both the TPJ and ACC (Grezes et al., 2004; Grezes et al., 2006). On the other hand, Drolet and colleagues (2011) found increased activation due to authentic stimuli during explicit perception of authenticity and argued that this may be due to the greater importance of authentic emotional content. A similar effect could be involved in this case, with authentic cues perceived as more important in emotion perception. In either case, cues and stimulus features can induce activation in parts of the TOM network, and while this does not necessarily translate into changes in explicit emotion recognition, in a more natural day-to-day context, such additional information may be integrated with the multitude of other contextual and multimodal information processed in parallel (Brück, Kreifelts, & Wildgruber, 2011; Rilliard, Shochi, Martin, Erickson, & Auberge, 2009; Scherer & Ellgring, 2007). Regenbogen et al. (2012) showed that changes in the information content from individual modalities can affect both ToM network activation and various stages of emotion recognition, including empathy. In other words, the information from various sources, with explicit authenticity representing one such component, can affect brain activation individually, with the potential to affect behavior through stimulus and modality integration, even if this effect was not enough for the present experimental task.

Play-acted cues clearly induced less differential activation overall, indicating that this type of cue did not increase attention to the stimuli. In fact, neither the TTG, the pSTS, nor the TPJ was shown to be activated in this contrast. As opposed to the activation correlated with authentic cues, these trials did not induce greater attention or increase mentalizing or perspective taking, in either the congruent or the incongruent condition. The difference between the influences of authentic and play-acted cues clearly shows that authentic cues influence perception more than do play-acted cues but also indicates how participants were able to ignore these cues and base their responses mostly on the perception of the stimuli.

The lack of TTG and pSTS modulation by authenticity during cued trials was an additional unexpected finding. Activation in the TTG and pSTS was induced by the stimuli spontaneously when no cue was given but was abolished when attention was directed to authenticity. While it remains unclear why, specifically, this activation is affected, top-down influences on early auditory activation are known (Friederici, 2012; Todorovic, van Ede, Maris, & de Lange, 2011), and since unexpected stimulus properties can have an

effect on BOLD response (Osnes, Hugdahl, Hjelmervik, & Specht, 2012), the differences between authentic and play-acted stimuli may have a greater effect when participants are not explicitly told about authenticity and are not expecting any contextual differences between stimuli.

Instruction differences would also explain why early auditory processing did not appear to be affected in Drolet et al. (2011), since participants were tasked with explicitly perceiving and categorizing stimuli as either authentic or play-acted. This would have caused participants to expect differences in context between stimuli, and the explicit nature of the task additionally influenced perception during emotion categorization in that study. In particular, this also appears to reverse the direction or effect of stimulus authenticity. While an explicit discrimination of authenticity puts more weight on authentic stimuli (also increasing activation for such stimuli), spontaneous reaction to authenticity appears to be up-regulated by play-acted stimuli. While this, and the potential influence of acoustical properties, would have to be corroborated by further research, the present results show that TPJ recruitment by differences in authenticity is spontaneous, while the mPFC and retrosplenial cortex were not affected, due to the lack of explicit authenticity recognition, as in Drolet et al. (2011).

Concluding remarks

While the interaction of emotion categorization and authenticity is already known from our previous work, the present results indicate that this effect can occur automatically and much earlier in cortical processing than previously shown. Without explicit perception or cueing, authenticity modulates early stimulus-related activity and, potentially, already induces deeper cognitive processing of the nature of the stimuli. The acoustical differences due to authenticity and the increased activation induced by play-acted stimuli indicate that, in a bottom-up manner, increased $f0$ contour (indicative of play-acted stimuli) induces spontaneous up-regulation in the BOLD response for regions involved in speech processing for both early auditory processing and later phonology-related activation and mentalizing. While an authentic cue leads participants to reassess their perception of the presented stimulus in cases of conflict, a play-acted cue is perceived as less relevant and reduces differential activation overall, and when cue and stimulus are perceived to be in conflict, stimulus perception simply overrides the cue, as indicated by the behavioral data. Therefore, while explicit authenticity information does not appear to have an impact on explicit behavior as measured in this study, modulation in brain activation indicates that it may nevertheless modulate our perception of emotion in real-life contexts.

References

Albert, J., López-Martín, S., & Carretié, L. (2010). Emotional context modulates response inhibition: Neural and behavioral data. *NeuroImage, 49*(1), 914–921. doi:10.1016/j.neuroimage.2009.08.045

Arnott, S. R., Binns, M. A., Grady, C. L., & Alain, C. (2004). Assessing the auditory dual-pathway model in humans. *NeuroImage, 22*(1), 401–408.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412. doi:10.1016/j.jml.2007.12.005

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*(2), 919–927.

Banse, R., & Scherer, K. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 70*(3), 614–636.

Barrett, L., & Kensinger, E. (2010). Context is routinely encoded during emotion perception. *Psychological Science*. doi:10.1177/0956797610363547

Baum, S., & Pell, M. (1999). The neural bases of prosody: Insights from lesion studies and neuroimaging. *Aphasiology, 13*(8), 581–608. doi:10.1080/026870399401957

Beaucousin, V., Lacheret, A., Turbelin, M., Morel, M., Mazoyer, B., & Tzourio-Mazoyer, N. (2007). FMRI study of emotional speech comprehension. *Cerebral Cortex, 17*(2), 339.

Blanchette, I., & Richards, A. (2010). The influence of affect on higher level cognition: A review of research on interpretation, judgement, decision making and reasoning. *Cognition & Emotion, 24*(4), 561. doi:10.1080/02699930903132496

Brück, C., Kreifelts, B., & Wildgruber, D. (2011). Emotional voices in context: A neurobiological model of multimodal affective information processing. *Physics of Life Reviews*. doi:10.1016/j.plrev.2011.10.002

Buchanan, T., Lutz, K., Mirzazade, S., Specht, K., Shah, N., Zilles, K., et al. (2000). Recognition of emotional prosody and verbal components of spoken language: An fMRI study. *Cognitive Brain Research, 9*(3), 227–238.

Burgoon, J. K., Blair, J. P., & Strom, R. E. (2008). Cognitive biases and nonverbal cue availability in detecting deception. *Human Communication Research, 34*(4), 572–599. doi:10.1111/j.1468-2958.2008.00333.x

Chung, Y. S., & Barch, D. M. (2011). The effect of emotional context on facial emotion ratings in schizophrenia. *Schizophrenia Research*. doi:10.1016/j.schres.2011.05.028

DePaulo, B. M. (2003). Cues to deception. *Psychological Bulletin, 129*(1), 74–118.

Drolet, M., Schubotz, R. I., & Fischer, J. (2011). Authenticity affects the recognition of emotions in speech: Behavioral and fMRI evidence. *Cognitive, Affective, & Behavioral Neuroscience*. doi:10.3758/s13415-011-0069-3

Ethofer, T. (2006). Impact of voice on emotional judgment of faces: An event–related fMRI study. *Human Brain Mapping, 27*(9), 707–714.

Friederici, A. D. (2012). The cortical language circuit: From auditory perception to sentence comprehension. *Trends in Cognitive Sciences*, (0). doi:10.1016/j.tics.2012.04.001

Friston, K., Fletcher, P., Josephs, O., Holmes, A., Rugg, M., & Turner, R. (1998). Event-related fMRI: Characterizing differential responses. *NeuroImage, 7*(1), 30–40.

Frith, C. (2008). Implicit and explicit processes in social cognition. *Neuron, 60*(3), 503–510.

Frith, U., & Frith, C. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society B: Biological Sciences, 358*(1431), 459–473. doi:10.1098/rstb.2002.1218

Goldstein, T. (2009). Psychological perspectives on acting. *Psychology of Aesthetics, Creativity, and the Arts, 3*(1), 6–9. doi:10.1037/a0014644

Goldstein, T. (2011). Correlations among social-cognitive skills in adolescents involved in acting or arts classes. *Mind, Brain, and Education, 5*(2), 97–103. doi:10.1111/j.1751-228X.2011.01115.x

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*(2), 601–608.

Grezes, J., Frith, C., & Passingham, R. (2004). Brain mechanisms for inferring deceit in the actions of others. *The Journal of Neuroscience, 24*(24), 5500–5505. doi:10.1523/JNEUROSCI.0219-04.2004

Hervé, P.-Y., Razafimandimby, A., Vigneau, M., Mazoyer, B., & Tzourio-Mazoyer, N. (2012). Disentangling the brain networks supporting affective speech comprehension. *NeuroImage, 61*(4), 1255–1267. doi:10.1016/j.neuroimage.2012.03.073

Hooker, C. I., Verosky, S. C., Germine, L. T., Knight, R. T., & D'Esposito, M. (2008). Mentalizing about emotion and its relationship to empathy. *Social Cognitive and Affective Neuroscience, 3*(3), 204.

Hornak, J., Rolls, E., & Wade, D. (1996). Face and voice expression identification in patients with emotional and behavioural changes following ventral frontal lobe damage. *Neuropsychologia, 34*(4), 247–261. doi:10.1016/0028-3932(95)00106-9

Jaywant, A., & Pell, M. D. (2011). Categorical processing of negative emotions from speech prosody. *Speech Communication, 54*(1), 1–10. doi:10.1016/j.specom.2011.05.011.

Jürgens, R., Hammerschmidt, K., & Fischer, J. (2011). Authentic and play-acted vocal emotion expressions reveal acoustic differences. *Frontiers in Emotion Science, 2,* 180. doi:10.3389/fpsyg.2011.00180

Kéri, S., Decety, J., Roland, P., & Gulyás, B. (2004). Feature uncertainty activates anterior cingulate cortex. *Human Brain Mapping, 21*(1), 26–33.

Kornbrot, D. E. (1978). Theoretical and empirical comparison of Luce's choice model and logistic Thurstone model of categorical judgment. *Perception & Psychophysics, 24,* 193–208. doi:10.3758/BF03206089

Kotz, S. A., Kalberlah, C., Bahlmann, J., Friederici, A. D., & Haynes, J. (2012). Predicting vocal emotion expressions from the human brain. *Human Brain Mapping.* doi:10.1002/hbm.22041

Kreifelts, B., Ethofer, T., Grodd, W., Erb, M., & Wildgruber, D. (2007). Audiovisual integration of emotional signals in voice and face: An event-related fMRI study. *NeuroImage, 37*(4), 1445–1456. doi:10.1016/j.neuroimage.2007.06.020

Linville, S. E. (1985). Acoustic characteristics of perceived versus actual vocal age in controlled phonation by adult females. *The Journal of the Acoustical Society of America, 78*(1), 40. doi:10.1121/1.392452

Lohmann, G., Müller, K., Bosch, V., Mentzel, H., Hessler, S., Chen, L., … Von Cramon, D. Y. (2001). Lipsia—a new software system for the evaluation of functional magnetic resonance images of the human brain. *Computerized Medical Imaging and Graphics*, 25(6), 449–457. doi:10.1016/S0895-6111(01)00008-8

Luce, R. (1959). *Individual choice behavior.* New York: Wiley.

MacDonald, A. W., Cohen, J. D., Stenger, V. A., & Carter, C. S. (2000). Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science, 288*(5472), 1835–1838. doi:10.1126/science.288.5472.1835

Mier, D., Lis, S., Neuthe, K., Sauer, C., Esslinger, C., Gallhofer, B., & Kirsch, P. (2010). The involvement of emotion recognition in affective theory of mind. *Psychophysiology*, no–no. doi:10.1111/j.1469-8986.2010.01031.x

Nelson, N. L., & Russell, J. A. (2011). Putting motion in emotion: Do dynamic presentations increase preschooler's recognition of emotion? *Cognitive Development, 26*(3), 248–259. doi:10.1016/j.cogdev.2011.06.001

Norris, D. (2000). Reduced power multislice MDEFT imaging. *Journal of Magnetic Resonance Imaging, 11*(4), 445–451.

O'Brien, M., Miner Weaver, J., Nelson, J. A., Calkins, S. D., Leerkes, E. M., & Marcovitch, S. (2011). Longitudinal associations between children's understanding of emotions and theory of mind. *Cognition & Emotion, 25,* 1074–1086. doi:10.1080/02699931.2010.518417

Osnes, B., Hugdahl, K., Hjelmervik, H., & Specht, K. (2012). Stimulus expectancy modulates inferior frontal gyrus and premotor cortex activity in auditory perception. *Brain and Language.* doi:10.1016/j.bandl.2012.02.002

Pell, M. (2005). Prosody–face interactions in emotional processing as revealed by the facial affect decision task. *Journal of Nonverbal Behavior, 29*(4), 193–215. doi:10.1007/s10919-005-7720-z

Pell, M., Monetta, L., Paulmann, S., & Kotz, S. (2009). Recognizing emotions in a foreign language. *Journal of Nonverbal Behavior, 33*(2), 107–120. doi:10.1007/s10919-008-0065-7

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *The Behavioral and Brain Sciences, 1*(04), 515–526. doi:10.1017/S0140525X00076512

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., & Striano, T. (2004). Young children know that trying is not pretending: A test of the "behaving-as-if" construal of children's early concept of pretense. *Developmental Psychology, 40*(3), 388–399. doi:10.1037/0012-1649.40.3.388

Redcay, E. (2008). The superior temporal sulcus performs a common function for social and speech perception: Implications for the emergence of autism. *Neuroscience and Biobehavioral Reviews, 32*(1), 123–142.

Regenbogen, C., Schneider, D. A., Gur, R. E., Schneider, F., Habel, U., & Kellermann, T. (2012). Multimodal human communication — targeting facial expressions, speech content and prosody. *NeuroImage, 60*(4), 2346–2356. doi:10.1016/j.neuroimage.2012.02.043

Rilliard, A., Shochi, T., Martin, J.-C., Erickson, D., & Auberge, V. (2009). Multimodal indices to Japanese and French prosodically expressed social affects. *Language and Speech, 52*(2–3), 223–243. doi:10.1177/0023830909103171

Samson, D., Apperly, I. A., Chiavarino, C., & Humphreys, G. W. (2004). Left temporoparietal junction is necessary for representing someone else's belief. *Nature Neuroscience, 7*(5), 499–500. doi:10.1038/nn1223

Saxe, R. (2006). Uniquely human social cognition. *Current Opinion in Neurobiology, 16*(2), 235–239.

Scherer, K. (1991). Vocal cues in emotion encoding and decoding. *Motivation and Emotion, 15*(2), 123–148.

Scherer, K., & Ellgring, H. (2007). Multimodal expression of emotion: Affect programs or componential appraisal patterns. *Emotion, 7*(1), 158–171.

Schirmer, A. (2010). Mark my words: Tone of voice changes affective word representations in memory. (A. Aleman, Ed.) *PLoS ONE, 5*(2), e9080. doi:10.1371/journal.pone.0009080

Schirmer, A., & Kotz, S. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences, 10*(1), 24–30. doi:10.1016/j.tics.2005.11.009

Schmidt, M. E., Pempek, T. A., Kirkorian, H. L., Lund, A. F., & Anderson, D. R. (2008). The effects of background television on

the toy play behavior of very young children. *Child Development, 79*(4), 1137–1151.

Schötz, S. (2007). Acoustic analysis of adult speaker age. *Speaker Classification I*, 88–107.

Sebastian, C. L., Fontaine, N. M. G., Bird, G., Blakemore, S.-J., Brito, S. A. D., McCrory, E. J. P., et al. (2012). Neural processing associated with cognitive and affective theory of mind in adolescents and adults. *Social Cognitive and Affective Neuroscience, 7*(1), 53–63. doi:10.1093/scan/nsr023

Shamay-Tsoory, S. G. (2011). The neural bases for empathy. *The Neuroscientist, 17*(1), 18–24.

Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain: 3-dimensional proportional system: An approach to cerebral imaging*. New York: Thieme.

Teufel, C., Fletcher, P., & Davis, G. (2010). Seeing other minds: Attributed mental states influence perception. *Trends in Cognitive Sciences, 14*(8), 376–382. doi:10.1016/j.tics.2010.05.005

Todorovic, A., Van Ede, F., Maris, E., & De Lange, F. P. (2011). Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: An MEG study. *The Journal of Neuroscience, 31*(25), 9118–9123. doi:10.1523/JNEUROSCI.1425-11.2011

Truett, A., Puce, A., & McCarthy, G. (2000). Social perception from visual cues: Role of the STS region. *Trends in Cognitive Sciences, 4*(7), 267–278.

Ugurbil, K., Garwood, M., Ellermann, J., Hendrich, K., Hinke, R., Hu, X., … Ogawa, S. (1993). Imaging at high magnetic fields: Initial experiences at 4T. *Magnetic Resonance Quarterly, 9*(4), 259–277.

Van Overwalle, F., & Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: A meta-analysis. *NeuroImage, 48*(3), 564–584. doi:10.1016/j.neuroimage.2009.06.009

Vigneau, M. (2006). Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing. *NeuroImage, 30*(4), 1414–1432.

Vogt, B. A. (2005). Pain and emotion interactions in subregions of the cingulate gyrus. *Nature Reviews Neuroscience, 6*(7), 533–544.

Vollm, B. A., Taylor, A. N.., Richardson, P., Corcoran, R., Stirling, J., McKie, S., … Elliott, R. (2006). Neuronal correlates of theory of mind and empathy: A functional magnetic resonance imaging study in a nonverbal task. *Neuroimage, 29*(1), 90–98.

Wildgruber, D., Pihan, H., Ackermann, H., Erb, M., & Grodd, W. (2002). Dynamic brain activation during processing of emotional intonation: Influence of acoustic parameters, emotional valence, and sex. *NeuroImage, 15*(4), 856–869.

Wildgruber, D., Riecker, A., Hertrich, I., Erb, M., Grodd, W., Ethofer, T., et al. (2005). Identification of emotional intonation evaluated by fMRI. *NeuroImage, 24*(4), 1233–1241.

Worsley, K., & Friston, K. (1995). Analysis of fMRI time-series revisited - again. *NeuroImage, 2*(3), 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*. doi:10.1002/hbm.21176