Out of the Spotlight: Neurophysiological Mechanisms Underlying the Processing of Unattended Vocal Expressions

Annett Schirmer[1], *Nicolas Escoffier*[1], *and Angela D.Friederici*[2]

[1]Department of Psychology, National University of Singapore
[2]Max Planck Institute for Human Cognitive and Brain Sciences
schirmer@nus.edu.sg

Abstract

The voice is a powerful medium for the communication of emotion. However, in conversation listeners typically attend to the content of a message rather than to the way it is communicated. Moreover, given the serial nature of speech processing, they focus on one talker at a time. Nevertheless, research on vocal processing indicates that listeners tend to be sensitive to emotional expressions from unattended sources. For example, unattended syllable sequences that rarely and unpredictably change in speaker prosody were found to elicit a larger mismatch negativity in the event-related potential (ERP) if syllables suddenly become emotional as compared to neutral. The present functional magnetic resonance imaging (fMRI) study aimed at elucidating the brain structures that mediate this effect. The observed evidence suggests that enhanced responses to unattended emotional as compared to neutral speaker prosody are mediated by activity in the middle frontal gyrus, the posterior superior temporal cortex, the anterior insula and the amygdala. Interestingly, however, listeners differ in their sensitivity to unattended and unexpected changes in speaker tone of voice. These differences appear to be linked to biological (e.g., estrogen) and cultural factors (e.g., social orientation) and affect how these individuals pursue and engage in social interactions.

1. Introduction

Sayings such as “Listen to your heart!” or “His anger clouded his thoughts.” reflect the popular belief that human behavior is at the mercy of two opposing forces: emotion and cognition. Rooted in ancient Greek philosophy, this belief has influenced whether and how scientists across past centuries approached the study of human emotion [5]. However, recent insights from neuroscience implicate overlapping neural networks in emotional and cognitive processing. Moreover, the discovery of close linkages and interactions has cast some doubt on the view that emotion and cognition are distinct and independent faculties of the human mind.

Past research on the processing emotional prosody has contributed to this change in how we think about the relationship between feelings and thoughts. For example, there is evidence that listeners recruit brain structures implicated in verbal processing and/or working memory when judging the emotional content of prosody [15, 27]. However, even if not intentionally encoding emotional prosody, listeners easily pick up on the emotion of their interaction partners and these emotions guide verbal processing [25, 24]. The processing of words or phrases presented in the context of incongruous emotional prosody is more effortful than the processing of verbal information that accords with speakers' tone of voice. For example, the measurement of event-related potentials (ERPs) through scalp electrodes placed on the listener’s head shows a larger negativity to positive words (e.g., success) spoken with an angry as compared to a happy voice [24]. Furthermore, functional magnetic resonance imaging (fMRI) revealed increased activity in the bilateral inferior frontal gyrus to words spoken with incongruous as compared to congruous emotional prosody [27].

Aside from modulating language processing, speaker emotion has also been shown to influence listener attention. Enhanced activation of the superior temporal sulcus (STS) to attended as compared to ignored sounds can also be seen as a function of emotion. That is, regardless of attentional focus, emotional prosody activates the right STS more strongly than neutral prosody [9]. Additional evidence for an influence of speaker emotion on listener attentional systems comes from ERPs. A study using an auditory oddball paradigm revealed a greater mismatch negativity (MMN) to auditory oddballs presented with an emotional as compared to neutral prosody [26, 23]. Given the presumed role of the MMN in auditory change detection [16], this work suggests that listeners are more sensitive to unattended change in speaker tone of voice if that change is emotional.

While the work on vocal emotional processing clearly demonstrates a link between emotion and cognition, it also suggests that this link may vary between individuals. Specifically, there is evidence suggesting that language processing is more susceptible to influences from speaker emotion in female as compared to male listeners [25, 24, 27]. Furthermore, female listeners are more likely than male listeners to show a larger mismatch negativity to emotional as compared to neutral prosody if these stimuli are presented outside of attentional focus. As these differences show for vocalizations but are non-significant for nonsocial auditory sounds [23], one may speculate that the social relevance of the auditory stimulus is critical in eliciting enhanced processing in females. Moreover, the observed processing enhancement may result from the influence of cultural and/or biological variables that modulate sensitivity to social emotional signals and the interaction between emotion and cognition in interpersonal communication.

Here we report an experiment that was aimed at elucidating one of these variables. The experiment utilized the auditory oddball paradigm reported above [26, 23]. Listeners were engaged in watching a silent, subtitled movie while passively listening to a sequence of spoken syllables that rarely and unpredictably changed in emotional prosody. Listener responses to these changes were measured using fMRI. This should reveal the brain structures that mediate the processing of vocal change and the discrimination between emotional and neutral tone of voice. Based on prior evidence we predicted that vocal change activates the auditory cortex [3, 33], whereas emotional discrimination was thought to recruit the STS and the amygdala [9, 21]. Following each experimental session, we measured listeners' social orientation or interest in social relationships with a questionnaire. We were interested in social orientation because women score higher than men on social orientation questionnaires [4]. Moreover, we speculated that listeners with high social orientation are inadvertently interested in social signals and may hence respond more strongly to these signals, especially if they are of emotional significance.

2. Methods

2.1. Participants

Fourteen female participants were recruited for this experiment. They had a mean age of 25.6 years (SDV 3.2) and had no known hearing impairments.

2.2. Material

A female speaker produced the syllable sequence "dada" with angry and neutral intonation. Angry and neutral syllables were equally long (angry/neutral: 557 ms) and loud (angry/neutral: 67 dB max, 56 dB mean) but differed with respect to fundamental frequency and other frequency formants.

2.3. Experimental Design

Angry and neutral syllables served as standards and deviants in two types of blocks. During one type of block, neutral syllables were presented with a high probability of 0.875 and angry syllables with a low probability of 0.125. During the other type of blocks, the probability of angry and neutral syllables was reversed. The syllable-onset to syllable-onset interval was 1.2 seconds. An experimental session comprised 4 blocks of each type with 200 stimuli each. Block types were presented in alternating order interspersed by a 20 second silent interval.

Sounds were presented over headphones and participants were instructed to ignore the sounds and to concentrate on a self-selected silent, subtitled movie. After the experiment participants completed a questionnaire assessing their social orientation [4].

2.4. MRI Scanning and Analysis

The data was collected on a 3-T MRI scanner. Twenty axial slices (19.2 cm FOV, 64 by 64 matrix, 4 mm thickness, 1 mm spacing), parallel to the AC-PC plane and covering the whole brain were acquired using a single shot, gradient recalled EPI sequence (TR 2000 ms, TE 30 ms, 90° flip angle). We recorded one functional run with 1044 repetitions. Prior to the functional run, a 20 slice anatomical T1-weighted MDEFT [34, 17] image (data matrix 256x256, TR 1.3 s, TE 10 ms) and a 20 slice T1-weighted EPI image with the same spatial orientation as the functional data were acquired.

LIPSIA [13] was used for data preprocessing. Differences in slice acquisition time were corrected with a cubic-spline interpolation. A temporal high-pass filter (1/700 Hz) was applied for baseline correction and a Gaussian filter (5.7mm) for spatial smoothing. To align the functional data slices onto a standard stereotactic space, a rigid linear registration was performed. The registration parameters were computed on the basis of the MDEFT and EPI-T1 scans to achieve an optimal match between these and an individual 3D reference data set, acquired during a previous session. The registration parameters and the 3D reference data set were then normalized to the Talairach stereotactic space [30]. To normalize the functional images into the same stereotactic space, a trilinear interpolation using the normalized registration parameters was applied. Furthermore, scans from each participant were subjected to a nonlinear normalization to a 3D reference data set [32].

The statistical evaluation was based on a least squares estimation using the general linear model for serially autocorrelated observations [6] and a two-stage, random effect analysis [12]. The design matrix comprised seven experimental regressors, four of which represented the 4 experimental conditions (i.e., emotional and neutral deviants and standards). Only 100 events were selected for each standard condition, by randomly picking one standard event from among the events that preceded a deviant. Unmodeled emotional and neutral standards, silent periods and movement parameters were added to the model as additional regressors. The data were filtered with a Gaussian kernel (4s FWHM). The model was convolved with a synthetic haemodynamic response function [8] and the fit between the data and the model was estimated. Effects were considered significant at $p < .0001$ or $Z > 3.1$.

3. Results

Table 1: *Effects of Social Orientation on the Deviant > Standard Contrast*

Anatomical Region	BA	Talairach Coordinates X	Y	Z
Precentral Gyrus	6	40	5	18
MFG/IFS	46	-35	44	12
MFG/IFS	46	43	41	9
Inferior Frontal Gyrus	45	46	20	18
Cuneus	18	19	-67	15
Cerebellum		34	-73	-21
Cerebellum		13	-49	-24

BA = Brodmann Area; MFG = Middle Frontal Gyrus; IFS = Inferior Frontal Sulcus

Table 2: *Emotional Deviant > Neutral Deviant*

Anatomical Region	BA	Talairach Coordinates X	Y	Z
Middle Frontal Gyrus*	10	34	50	15
Posterior STS	42/39	-41	-35	12
Anterior Insula	10	-29	34	3
Amygdala/Hippocampus*		31	-14	-15

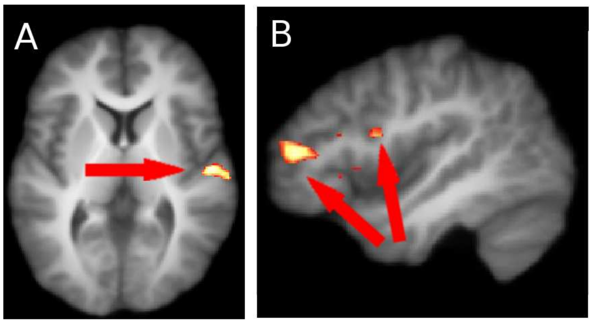
BA = Brodmann Area; *p < .005; STS = Superior Temporal Sulcus

Change in speaker prosody elicited increased activity in the auditory cortex lateralized to the right hemisphere. A second level regression analysis with social orientation as a predictor indicated that social orientation modulated activity in a number of structures. Specifically, with increased social orientation, deviants elicited greater activation than standards in the prefrontal cortex lateralized to the right hemisphere as well as in the cuneus and the cerebellum (see Table1, Figure 1).

A direct contrast between emotional and neutral deviants revealed activity in the left posterior STS and the left anterior insula. With a slightly lower significance threshold ($p > .005$), additional activation was observed in the right middle frontal gyrus (MFG) and the right amygdala.

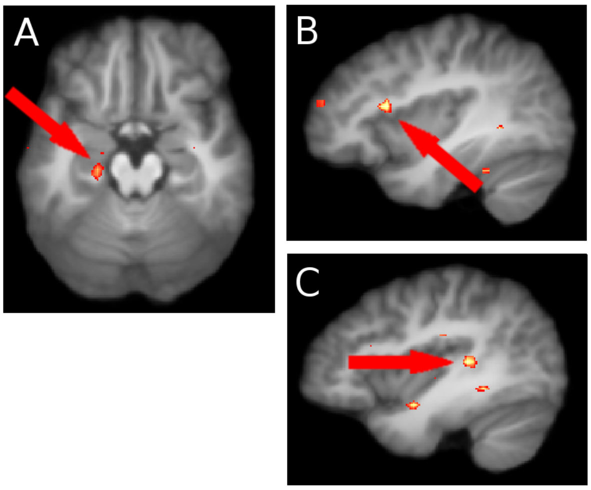
An additional whole brain regression analysis testing for a modulation of the emotional vs. neutral deviant contrast by social orientation failed to reach significance.

Figure 1: *Effects of Vocal Change and Social Orientation*.



(A) Activity in Heschl's Gyrus observed in a direct contrast between deviants and standards. (B) Activity in the prefrontal cortext observed in a second level analysis investigating the relationship between social orientation and processing differences between deviants and standards.

Figure 2: *Effects of Emotional Significance*.



Illustrated is a direct contrast between emotional and neutral deviants. This activated the amygdala (A), the anterior insula (B) and the posterior STS (C).

4. Discussion

The present study employed a passive auditory oddball paradigm to unravel the neurophysiological mechanisms that allow listeners to discriminate between emotional and neutral vocalizations presented outside the focus of attention. Moreover, we were interested in whether these mechanisms would differ as a function of a listener's social orientation. We found that listeners in the present study showed increased responses in the auditory cortex to rare changes in speaker prosody regardless of whether these changes were neutral or emotional. The emotionality of prosody seemed to affect activity in other structures such as the right MFG, left posterior STS, left anterior insula and right amygdala.

When interpreting these findings, it is important to consider prior research on auditory change detection. This work indicated that auditory deviants that differ from their acoustic context in one or more physical attributes activate regions in Heschl's gyrus and the STG [3, 33]. Based on this evidence and evidence from source localization studies that located MMN generators in these regions [11], it has been proposed that they represent the seat of the sensory memory comparison that mediates auditory change detection. In line with this, we observed that rare vocal change, which elicits an MMN in the ERP [26, 23], also activates the auditory cortex. That this activation fails to differ as a function of emotional significance suggests that basic aspects of auditory change detection respond primarily to acoustic rather than to emotional stimulus properties. Moreover, the emotional significance of vocal change appears to recruit processing components that follow basic sensory encoding.

Based on the present results these processing components can be identified as subserving vocal and emotional analysis. The right MFG region that we observed in the contrast between emotional and neutral change matches findings from previous fMRI studies [7, 37, 38]. In these studies, emotional prosodic judgments were contrasted with another type of task, isolating processes involved in the conscious evaluation of emotional significance. However, as the right MFG has also been reported in the emotional evaluation of nonvocal stimuli [e.g., 29] it may play a more general role in emotional processing and represent a cognitive appraisal component that operates across modalities.

The present study also revealed the left posterior STS as an area that is more strongly activated by emotional as compared to neutral prosodic change. Given previous evidence of voice selective regions along the STS [1] one may speculate that the present findings reflect enhanced vocal processing of emotional as compared to neutral stimuli. However, voice selective regions have been reported to be located more anteriorly than the present activation. Moreover, vocal emotional discrimination in the STS has been reported to be right lateralized [9]. Based on these discrepancies, we would like to argue that the STS activation observed here is not voice specific. This assumption is in line with the presumed functions of the posterior STS. Being located along the auditory "where" stream, the posterior STS may contribute to spatial auditory processing [35, 30]. Vocal change presented over headphones was spatially distinct from participants' attentional focus directed at the screen in front of them. Hence, emotionally significant change may have triggered spatial auditory processing associated with orienting towards the sound source. However, as the posterior STS is anatomically and functionally heterogeneous, alternative interpretations are possible. Specifically, the presence of auditory and visual input in this area makes it a candidate for audio-visual integration [28]. Listeners may engage in such integration to a greater extent when presented with emotional as compared to neutral information in one or more sensory modalities.

The emotional enhancement of cognitive appraisal and audio-visual processing observed here were accompanied by activity in the left anterior insula and, to a smaller degree, in the right amygdala. Both regions have been implicated in emotional processing across stimulus modalities. The insula plays a role in olfactory, gustatory and autonomic processing, and thus appears to be a good candidate for the integration of sensory and emotional information and the mediation of bodily feelings [2]. Moreover, its consistent involvement in the processing of social emotional cues led researchers to propose that it may allow perceivers to resonate emotions displayed by senders [10, 36]. The amygala is also involved in mediating bodily responses to emotional stimuli [19]. Additionally, it serves to upregulate the cognitive processing of an emotion eliciting event. As such the amygdala has been shown to enhance sensory and attentional processes as well as memory storage [19]. Moreover, as amygdala activity may occur independently from cortical activity and cognitive appraisal it has been proposed to reflect a highly automatic mechanism that is capable of appraising the relevance of a stimulus based on crude sensory information [20]. The present and previous vocal emotional processing studies [20] suggest that this mechanism is also activated in response to emotionally significant vocal cues.

There is evidence that judgments of and brain activity elicited to emotional signals may differ between individuals [25, 24, 27, 26, 23]. For example, women have been shown to be more responsive to emotionally salient stimuli than men particularly when these stimuli are socially relevant [23]. Based on concurrent differences between men and women in social orientation, we speculated that social orientation may account for such gender effects. Moreover, we predicted that individuals with high social orientation would be more sensitive to vocal emotional change as this change should be more relevant to them than to individuals with low social orientation. The present study failed to support this proposition. While social orientation enhanced sensitivity to unattended vocal change in general, no specific enhancement was observed when vocal change was emotional. The general effect of social orientation on activity in the prefrontal cortex, cuneus and the cerebellum is nevertheless interesting as it matches observations from visual and memory search tasks [14]. Combined activity in these structures has been proposed to reflect a general search mechanism that is common across different tasks. The present results suggest that such a general search mechanism may also be elicited in highly socially oriented listeners in response to sudden changes in the tone of voice of an unattended speaker. Moreover, these changes may trigger an orienting response that is aimed at locating or identifying the person that produced the unexpected vocalization.

That differences in the search for or orientation to vocal emotional and neutral change are independent of a listener's social orientation may be due to ceiling effects. Social orientation may enhance both the detection of neutral and emotional change such that responses to emotional change are already at ceiling. Alternatively, however, other mechanisms may account for enhanced responses to emotional as compared to neutral vocalizations. For example, there is evidence that estrogen modulates the accuracy of facial emotion recognition [18]. Likewise, vocal emotional processing has been shown to correlate with estrogen. Regardless of gender, differences in MMN amplitude between emotional and neutral changes in tone of voice were larger in listeners with high as compared to low estrogen [22]. Aside from estrogen, other biological or cultural variables may shape an individual's response to emotional signals. However, only with the recently renewed interest in interindividual differences and genome-behavior contingencies will these variables be discovered.

5. Conclusions

Navigating human's complex social life requires a sophisticated cognitive system that is sensitive to the emotional significance of social signals. Such sensitivity increases the likelihood that important information is both attended to and stored in memory. The present study shows that the tone of voice of unattended speakers may be more likely to capture listener attention if it is emotional as compared to neutral. Moreover, activity in the MFG, posterior STS, insula and amygdala seems at the core of this effect. These structures, subserving emotion recognition, audio-visual and/or emotional processing, facilitate the processing of emotional vocalizations and thus their chance of influencing subsequent cognition and behavior. While there is evidence for interindividual differences in the sensitivity to emotional prosody, the variables that drive these differences are still poorly understood. Social orientation, the variable investigated here, clearly enhances responses to vocal change. However, this enhancement appears to be non-specific with respect to the emotional significance of vocal change. The variables that influence this latter, emotional aspect of vocal processing have to be identified in future studies.

6. References

[1] Belin, P., Fecteau, S., Bedard, C. 2004. Thinking the voice: neural correlates of voice perception. Trends in Cognitive Sciences, 8, 129-135.

[2] Critchley, H.D., Wiens, S., Rotshtein, P., Őhman, A., Dolan, R.J. 2004. Neural systems supporting interoceptive awareness. Nature Neuroscience, 7, 189-195.

[3] Doeller, C.F., Opitz, B., Mecklinger, A., Krick, C., Reith, W., Schroeger, E. 2003. Prefrontal cortex involvement in preattentive auditory deviance detection: neuroimaging and electrophysiological evidence. NeuroImage, 20, 1270-1282.

[4] Fahrenberg, J., Hampel, R., Selg, H. 1994. Das Freiburger Persoenlichkeitsinventar FPI: Revidierte Fassung FPI-R and teilweise geaenderte Fassung FPI-A1. Hogrefe: Geottingen.

[5] Finger, S., 1994. Origins of Neuroscience. New York: Oxford University Press.

[6] Friston, K.J., 1994. Statistical parametric maps in functional imaging: a general linear approach. Human Brain Mapping, 2, 189–210.

[7] Gandour, J., Wong, D., Dzemidzic, M., Lowe, M., Tong, Y., Li, X. 2003. A cross-linguistic fMRI study of perception of intonation and emotion in Chinese. Human Brain Mapping, 18, 149–157.

[8] Glover, G.H. 1999. Deconvolution of impulse response in event-related BOLD fMRI. NeuroImage, 9, 416-429.

[9] Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M.L., Scherer, K.R., Vuilleumier, P. 2005. The voices of wrath: brain responses to angry prosody in meaningless speech. Nature Neuroscience, 8, 145-146.

[10] Grosbras, M.H., Paus, T. 2005. Brain networks involved in viewing angry hands and faces. Cerebral Cortex, 16, 1087-1096.

[11] Ha, K.S., Youn, T., Kong, S.W., Park, H.J., Ha, T.H., Kim, M.S., Kwon, J.S. 2003. Optimized individual mismatch negativity source localization using a realistic head model and the Talairach coordinate system. Brain Topography, 15, 233-8.

[12] Holmes, A.P., Friston, K.J. 1998. Generalisability, random effects and population inference. NeuroImage, 7, S754.

[13] Lohmann, G., Muller, K., Bosch, V., Mentzel, H., Hessler, S., Chen, L., Zysset, S., 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, 449- 457.

[14] Makino, Y., Yokosawa, K., Takeda, Y., Kumada, T. 2004. Visual search and memory search engage extensive overlapping cerebral cortices: an fMRI study. NeuroImage, 23, 525-533.

[15] Mitchell, R.L. 2007. FMRI delineation of working memory for emotional prosody in the brain: Commonalities with the lexico-semantic emotion network. NeuroImage, 36, 1015-1025.

[16] Näätänen, R., Jacobsen, T., Winkler, I., 2005. Memory based of afferent processes in mismatch negativity (MMN): A review of evidence. Psychophysiology. 42, 25-32.

[17] Norris, D.G., 2000. Reduced power multi-slice MDEFT imaging. Journal of Magnetic Resonance Imaging, 11, 445-451.

[18] Pearson, R., Lewis, M.B., 2005. Fear recognition across the menstrual cycle. Hormones and Behavior, 47, 267-271.

[19] Phelps, E.A. 2006. Emotion and cognition: insights from studies of the human amygdala. Annual Review of Psychology, 57, 27-53.

[20] Sander, D., Grafman, J., & Zalla, T. (2003). The human amygdala: an evolved system for relevance detection. Reviews in the Neurosciences, 14, 303-316.

[21] Sander, D., Grandjean, D., Pourtois, G., Schwartz, S., Seghier, M.L., Scherer, K.R., Vuilleumier, P. 2005. Emotion and attention interactions in social cognition: brain regions involved in processing anger prosody. NeuroImage, 28, 848-58.

[22] Schirmer, A., Escoffier, N., Li, Q.Y., Li, H., Strafford-Wilson, J., Li, W.I. 2007. What grabs his attention but not hers: Estrogen correlates with neurophysiological measures of vocal change detection. Under review.

[23] Schirmer, A., Escoffier, N., & Simpson, E. (2007). Listen up! Processing of intensity change differs for vocal and nonvocal sounds. *Brain Research, 1176*, 103-112.

[24] Schirmer, A., Kotz, S.A. 2003. ERP evidence for a gender specific Stroop effect in emotional speech. Journal of Cognitive Neuroscience, 15, 1135-1148.

[25] Schirmer, A., Kotz, S.A., Friederici, A.D. 2002. Sex differentiates the role of emotional prosody during word processing. Cognitive Brain Research, 14, 228-233.

[26] Schirmer, A., Striano, T., Friederici, A.D. 2005. Sex differences in the pre-attentive processing of vocal emotional expressions. Neuroreport, 16, 635-639.

[27] Schirmer, A., Zysset, S., Kotz, S.A., von Cramon, D.Y. 2004. Gender differences in the activation of inferior frontal cortex during emotional speech perception. NeuroImage, 21, 1114-1123.

[28] Scott, S.K., Johnsrude, I.S. 2003. The neuroanatomical and functional organization of speech perception. Trends in Neurosciences, 26, 100-107.

[29] Stark, R., Zimmermann, M., Kagerer, S., Schienle, A., Walter, B., Weygandt, M., Vaitl, D. 2007. Hemodynamic brain correlates of fear and disgust ratings. NeuroImage, 37, 663-673.

[30] Tata, M.S., Ward, L.M. 2005. Early phase of mismatch negativity is localized to a posterior "where" auditory pathway. Experimental Brain Research, 167, 481-486.

[31] Talairach, J., Tournoux, P. 1988. Coplanar stereotaxic atlas of the human brain. New York, NY: Thieme.

[32] Thirion, J.-P. 1998. Image matching as a diffusion process: an analogy with Maxwell's demons. Medical Image Analysis, 2, 243-260.

[33] Tse, C.Y., Tien, K.R., Penney, T.B. 2006. Event-related optical imaging reveals the temporal dynamics of right temporal and frontal cortex activation in pre-attentive change detection. NeuroImage, 29, 314-320.

[34] Ugurbil, K., Garwood, M., Ellermann, J., Hendrich, K., Hinke, R., Hu, X., Kim, S.-G., Menon, R., Merkle, H., Ogawa, S., Salmi, R. 1993. Imaging at high magnetic fields: initial experiences at 4T. Magnetic Resonance Quarterly, 9, 259-277.

[35] Warren, J.E., Wise, R.J., Warren, J.D. 2005. Sounds do-able: auditory motor transformations and the posterior temporal plane. Trends in Neurosciences, 28, 636-643.

[36] Wicker, B., Keysers, C., Plailly, J., Royet, J.P., Gallese, V., Rizzolatti, G. 2003. Both of us disgusted in My insula: the common neural basis of seeing and feeling disgust. Neuron, 40, 655-664.

[37] Wildgruber, D., Hertrich, I., Riecker, A., Erb, M., Anders, S., Grodd, W., Ackermann, H. 2004. Distinct frontal regions subserve evaluation of linguistic and emotional aspects of speech intonation. Cerebral Cortex, 14, 1384–1389.

[38] Wildgruber, D., I., Riecker, A., Hertrich, I., Erb, M., Grodd, W., Ethofer, T., Ackermann, H. 2005. Identification of emotional intonation evaluated by fMRI. NeuroImage, 24, 1233–1241.