ISCA Archive
http://www.isca-speech.org/archive

Speech Prosody 2006
Dresden, Germany
May 2-5, 2006

efMRI Evidence for Implicit Emotional Prosodic Processing

Sonja A. Kotz, Silke Paulmann & Tim Raettig

Max Planck Institute of Human Cognitive and Brain Sciences
Leipzig, Germany
`kotz@cbs.mpg.de`

10.21437/SpeechProsody.2006-216

Abstract

The current efMRI experiment investigated the potential right hemisphere dominance of emotional prosodic processing under implicit task demands. Participants evaluated the relative tonal height (high, medium, low) of intelligible and unintelligible sentences spoken by a trained female speaker of German with three prosodic contours: happy, angry, and neutral. The results confirm the activation of a bilateral fronto-striato-temporal network with no clear right hemispheric preference for emotional prosodic processing. The data suggest that (1) task demands do not significantly alter lateralization of function in the current context, and (2) fronto-striatal brain areas engage during implicit processing of emotional prosody, thus do no seem to be task specific.

1. Introduction

Emotional prosody as a non-verbal carrier of a speaker's emotive state and intention plays a crucial role during social communicative interaction. While there is a long established notion that emotional prosody may be a right hemisphere function [1, 2, 3], so far clinical and neuroimaging evidence have resulted in rather opposing views on the neural network supporting this function [4]. While divergent patient selection and variable times of measurement have been noted as critical factors contributing to inconsistent clinical results, factors such as task demands [5], design [6], and the lack of differentiation of emotional prosody as a multi-step process [7] may have influenced divergent neuroimaging evidence. Here we focussed on task demands, as most previous neuroimaging studies of emotional prosodic processing have used explicit tasks (i.e., categorization, identification, recognition). We investigated emotional prosody with an implicit task, that is, the task was not directed towards emotional prosody per se, in an event-related functional magnetic resonance imaging (efMRI) experiment. The motivation for the study was two-fold: (1) to find out if implicit processing of emotional prosody primarily engages the right hemisphere, and (2) to explore whether fronto-striatal areas are functionally specific for emotive evaluation [8, 9, 10].

2. Methods

2.1. Participants

Twenty-two right-handed native speakers of German (12 female) participated in the experiment. The average age of the participants was 24.25 years (SD: 1.84). All participants had normal hearing and no neurological condition at the time of measurement. Participants were paid a financial compensation after the experiment.

2.2. Material

A trained female speaker of German spoke semantically neutral sentences in three prosodic contours with either a happy, angry or neutral intonation forming the intelligible speech condition. All sentence were additionally filtered (PURR filter, [11]) rendering the sentences unintelligible, as the speech signal did not carry segmental or lexical information after filtering. However, the acoustic parameters were kept constant in both intelligible and unintelligible sentences.

2.3. Procedure and Task

Outside the scanner participants trained to evaluate the relative tonal height (high, medium, low) of emotional contours (six intelligible and six unintelligible emotional sentences with two sentences each spoken with happy, angry or neutral intonation; training sentences were not used in the experimental run) with response feedback after pressing one of three buttons for the respective answer. On average participants responded above 80% in all conditions after two training runs. In the scanner participants listened to 36 sentences per condition (high=happy, low=angry, medium=neutral) plus 40 null trials to enhance the variability in stimulus presentation resulting in a total number of 256 trials (108 intelligible sentences, 108 unintelligible sentences, 40 null trials). Intelligible and unintelligible sentences were presented in pseudorandomized order. After stimulus presentation participants were asked to evaluate the relative tonal height (high, medium, low) of intelligible and unintelligible emotional sentences spoken by a trained female speaker of German on an fMRI compatible three-button response pad (response order counterbalanced across participants). Participants did not receive response feedback in the scanner. The relation between image acquisition and trial presentation was systematically varied (0 to 3000 ms in 500 ms increments) so that numerous time points along the response were sampled [12] with an average inter-stimulus interval of 7 seconds, allowing the fMRI signal to decrease adequately between trials. Due to short duration of sentences (~ 1.5 seconds) the design was suitable for event-related analysis. The duration of the functional part of the study was approximately 30 minutes.

2.4. Data Acquisition

MRI data were collected at 3T using a Bruker 30/100 Medspec system (Bruker Medizintechnik GmbH, Ettlingen, Germany). The standard bird cage head coil was used. Before MRI data acquisition, field homogeneity was adjusted by means of 'global shimming' for each subject. Then, scout spin echo sagittal scans were collected to define the anterior and posterior commissures on a midline sagittal section. For each participant, structural and functional (echo-planar) images were obtained from 18 axial slices parallel (4 mm thickness, 1 mm inter-slice gap, 64x64 with a FOV of 19.2 mm, 3x3 mm in-plane resolution) to the plane intersecting the anterior and posterior commissures (AC-PC plane). After defining the slices' position a set of two-dimensional T1 weighted anatomical images (MDEFT sequence: TE 20 ms, TR 3500 ms, in-plane resolution 0.325 mm^2) were collected in plane with the echo-planar images to align the functional images to the 3D-images. A gradient-echo EPI sequence was used with a TE 30ms, flip angle 90 degrees, TR 3500 ms. In a separate session high resolution whole-head 3D MDEFT brain scans (128 sagittal slices, 1.5mm thickness, FOV 25.0x25.0x19.2cm, data matrix of 256x256 voxels) were acquired additionally for reasons of improved localization [13, 14].

2.5. Data Analysis

Behavioural results, calculated as percent correct responses, were statistically evaluated by means of a repeated measures ANOVA with two within-subjects factors: intelligibility (intelligible/unintelligible) and condition (high, medium, low). fMRI data were analysed with the LIPSIA software package [15]. Data preparation proceeded as follows: slice-wise motion correction (time step 50 as reference), cubic-spline interpolation in time (to correct for fMRI slice acquisition sequence); baseline correction; spatially smoothing using a Gaussian kernel with FWHM 5.65 mm. To align the functional data slices onto a 3D stereotactic coordinate reference system, a rigid linear registration with six degrees of freedom (3 rotational, 3 translational) was performed. The rotational and translational parameters were acquired on the basis of the 2-dimensional MDEFT volume to achieve an optimal match between these slices and the individual 3D reference dataset. The resulting parameters, scaled to standard size, were then used to transform the functional slices using trilinear interpolation.

Statistical evaluation was based on a least-square estimation using the general linear model for serially auto-correlated observations [16, 17]. First, for each subject, statistical parametric maps (SPM) were generated. The design matrix was generated with the standard hemodynamic response function considering a response delay of 6 seconds and its first and second derivative. The model includes an estimate of temporal autocorrelation. The effective degrees of freedom were estimated as described by Worsley and Friston [18]. Thereafter, contrast maps, (i.e., estimates of the raw-score differences of the beta coefficients between specified conditions), were generated for each subject. As the individual functional datasets were all aligned to the same stereotactic reference space, a group analysis was subsequently performed. A one-sample t-test of contrast maps across subjects was computed to indicate whether observed differences between conditions were significantly different from zero as suggested by Holmes and Friston [19]. Obtained t-values were subsequently transformed into z-values giving an SPM Z for each subject and condition. Voxels exceeding the threshold $|Z = 3.09|$, corresponding $P < 10^{-3}$ are reported as significant results.

3. Results

3.1. Behavioural results

Statistical analyses revealed a main effect of intelligibility (F (1,21) = 20.63, *p* <. 0002) indicating that participants evaluated the relative tonal height of intelligible sentences (mean: 86.5% (SD: 14.2 %) with greater confidence than for unintelligible sentences (mean: 78% (SD: 16.2%). A second main effect of condition (F (2,42) = 7.96, *p* < .001) showed that participants best evaluated medium tonal height (mean: 86% (SD: 13.6%), followed by high tonal height (mean: 83.5% (SD: 14.1%) and low tonal height (mean: 75.1% (SD: 17.9%). The interaction of intelligibility by condition approached significance (*p* = 0.1).

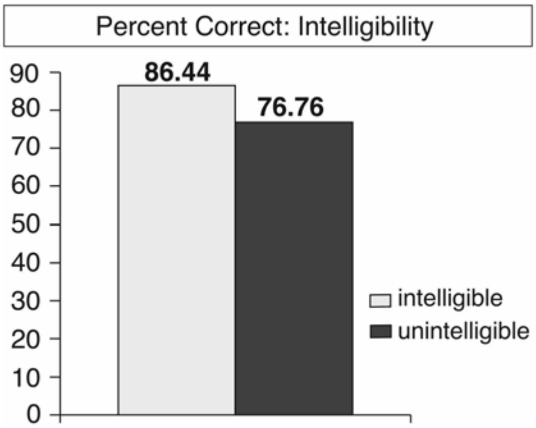


Figure 1*: Illustration of the behavioural response pattern of relative tonal height for intelligible and unintelligible sentences.*

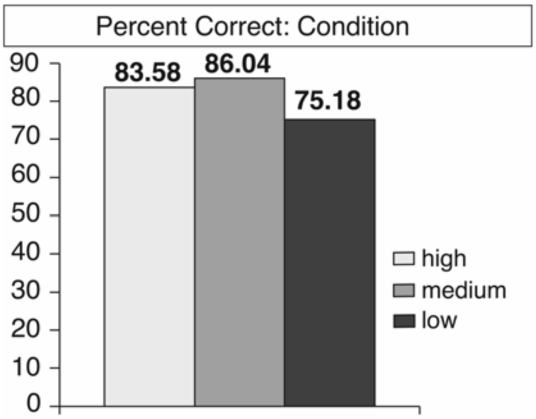


Figure 2: *Illustration of the behavioural response pattern of high, medium, and low prosodic contours.*

3.2. efMRI results

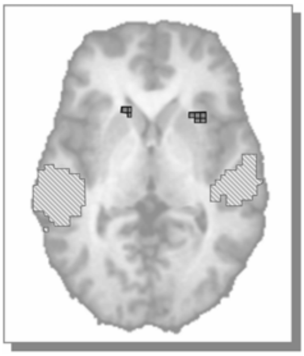
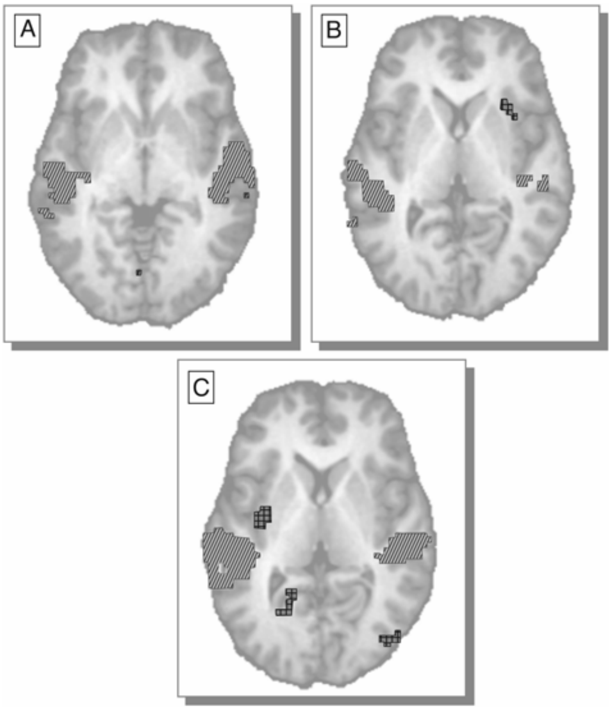


Figure 3: *Intelligibility effect in an axial view (z=2). Displayed is a bilateral temporal activation pattern for intelligible sentences (light grey striped) and bilateral striato-insular*

activation for unintelligible sentences (dark grey). Functional activation was thresholded at |Z = 3.09|, uncorrected.



Figures 4: *Condition effect in an axial view. Figure 4a (z=5) displays bilateral activation along the STG/STS for low contours (light grey=intelligible sentences), Figure 4b (z=5) left > right STG/STS activation for high contours (light grey=intelligible sentences) and right striato-insular activation for high contours (dark grey=unintelligible), and Figure 4c bilateral STG/STS activation for medium contours (light grey=intelligible sentences) as well as mainly left putaminal/insular activation for medium contours (dark grey=unintelligible). Functional activation was thresholded at |Z = 3.09|, uncorrected.*

4. Discussion/Conclusion

The current experiment set out to investigate two questions: (1) Does implicit processing of emotional prosodic information lead to a right hemispheric activation preference? and (2) Is previously reported fronto-striatal activation task specific rather than emotion specific? The present data clearly show that implicit processing of emotional prosody does not favour right hemisphere activation, but rather engages a bilateral activation of STG/STS for intelligible sentences and a bilateral striatal patterns for unintelligible sentences. Latter activation demonstrates that the striatum plays a crucial role in emotional prosodic processing that does not seem to be task dependent (see comparable results under explicit task demands in [9]). The current results are in line with previous reports that processing of prosodic information is supported by a bilateral brain network [9, 10, 20].

5. References

[1] Blonder, LX, Bowers, D, Heilman, KM (1991). The role of the right hemisphere in emotional communication. Brain 114: 1115-1127.

[2] Borod, JC (1993). Cerebral mechanisms underlying facial, prosodic, and lexical emotional expression: a review of neuropsychological studies and methodological issues. Neuropsychology 7: 445-463.

[3] Starkstein, E, Federoff, JP, et al. (1994). Neuropsychological and neuroradiological correlates of emotional prosody comprehension. Neurology 44: 515-522.

[4] Pell, MD, Baum, SR (1997). The ability to perceive and comprehend intonation in linguistic and affective contexts by brain-damaged adults. Brain Lang. 57: 80-89.

[5] Pell, MD (2006). Cerebral mechanisms for understanding emotional prosody in speech. Brain Lang. 96: 221-234.

[6] Kotz, SA, Meyer, M, Paulmann, S (in press). Impact of study design on lateralization of emotional prosody. Prog in Brain Res.

[7] Schirmer, A, Kotz, SA (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. TICS 10: 24-30.

[8] Cancelliere, A, Kertesz, A (1990). Lesion localization in acquired deficits of emotional expression and comprehension. Brain Cogn. 13: 133-147.

[9] Kotz, S. A., Meyer, M., et al. (2003). On the lateralization of emotional prosody: An event-related functional MR investigation. Brain Lang. 86: 366-376.

[10] Pell MD, Leonard, CL (2003). Processing emotional tone from speech in Parkinson's disease: a role for the basal ganglia. Cogn Affect Behav Neurosci. 3: 275-288.

[11] Sonntag, GP, & Portele, T (1998). PURR – a method for prosody evaluation and investigation. Journal of Comp Speech Lang. 12: 437-451.

[12] Burock MA, Buckner RL, et al. (1998). Randomized event-related experimental designs allow for extremely rapid presentation rates using functional MRI. NeuroReport 9:3735-3739.

[13] Norris DG (2000). Reduced power multi-slice MDEFT imaging. Magn Reson Imaging 11:445-451.

[14] Ugurbil K, Garwood ME, et al. (1993). Magnetic fields: Initial experiences at 4T. Magn Reson Q 9:259.

[15] Lohmann G, Müller K, et al. (2001). Lipsia - a new software system for the evaluation of functional magnetic resonance images of the human brain. Comput Med Imaging Graph 25:449-457.

[16] Friston KJ (1994), Statistical parametric maps in functional imaging: a general linear approach. Hum Brain Mapp 2:189-210.

[17] Zarahn E, Aguirre G, D'Esposito M (1997). Empirical analysis of BOLD-fMRI statistics. I. Spatially smoothed data collected under null-hypothesis and experimental conditions. NeuroImage 5:179-197.

[18] Worsley KJ, Friston KJ (1995). Analysis of fMRI time series revisited - again. NeuroImage 2:173-181.

[19] Holmes AP, Friston KJ (1998). Generalisability, random effects, and population inference. NeuroImage 7:754.

[20] Gandour, JY, Tong, Y (2004). Hemispheric roles in the perception of speech prosody. Neuroimage 23: 344-357.