COMMENTARY

A cerebral network model of speech prosody comprehension

DIRK WILDGRUBER1, THOMAS ETHOFER1, DIDIER GRANDJEAN2, & BENJAMIN KREIFELTS1

1University of Tuebingen, Germany, and 2University of Geneva, Switzerland

Abstract

Comprehension of information conveyed by the tone of voice is highly important for successful social interactions (Grandjean et al., 2006). Based on lesion data, a superiority of the right hemisphere for cerebral processing of speech prosody has been assumed. According to an early neuroanatomical model, prosodic information is encoded within distinct right-sided perisylvian regions which are organized in complete analogy to the left-sided language areas (Ross, 1981). While the majority of lesion studies are in line with the assumption that the right temporal cortex is highly important for the comprehension of speech melody (Adolphs et al., 2001; Borod et al., 2002; Heilman et al., 1984), some studies indicate a widespread network of partially bilateral cerebral regions to contribute to prosody processing including the frontal cortex (Adolphs et al., 2002; Hornak et al., 2003; Rolls, 1999) and the basal ganglia (Cancelliere & Kertesz, 1990; Pell & Leonard, 2003). More recently, functional imaging experiments have helped to differentiate specific functions of distinct brain areas contributing to recognition of speech prosody (Ackermann et al., 2004; Schirmer & Kotz, 2006; Wildgruber et al., 2006). Observations in healthy subjects indicate a strong association of cerebral responses and acoustic voice properties in some regions (stimulus-driven effects), whereas other areas show modulation of activation linked to the focusing of attention to specific task components (task-dependent effects). Here we present a refined model of prosody processing and cross-modal integration of emotional signals from face and voice which differentiates successive steps of cerebral processing involving auditory analysis and multimodal integration of communicative signals within the temporal cortex and evaluative judgements within the frontal lobes.

Keywords: Prosody, adults, neuropsychology.

Stimulus-driven effects

Enhanced activation within the middle section of the superior temporal cortex (mid-STC) has been observed in response to human voices as compared to other acoustic signals (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000). Regarding emotional prosody, Grandjean and collaborators (Grandjean et al., 2005) have demonstrated increasing responses within this region related to anger prosody independent of the participant's spatial attention during a dichotic listening task. In another fMRI study, participants either rated the emotional valence of verbal content or the emotional valence of speech prosody. Independent of the task, enhanced activation within the mid-STC was associated with increasing intensity of emotional prosody (Ethofer et al., 2006a). These results concur with neuroimaging findings obtained during passive listening to adjectives and substantive words with neutral word content, spoken in five different emotional intonations (Wiethoff et al., 2008). All emotional categories (happy, fearful, angry and alluring) induced stronger responses within the right mid-STC than neutral stimuli. These responses were significantly correlated with several acoustic parameters (stimulus duration, mean intensity, mean pitch and pitch variability). Separate simple regression analyses revealed that none of the parameters alone could explain the observed activation pattern. Evaluating the conjoint effect of these acoustic parameters in a multiple regression analysis, however, sufficiently explained the increase of responses within the right mid-STC. Therefore, an important contribution of this area to the integration of several acoustic parameters into an emotional percept has been suggested (Wiethoff et al., 2008). Moreover, an analysis of interaction effects between the gender of speaker and listener revealed a cross-gender interaction with increasing responses to the voice of the opposite sex in male and female subjects. This effect was confined to an alluring tone of speech in behavioural data as well as hemodynamic responses within the mid-STC. The response pattern of the mid-STC, thus, indicates a particular sensitivity to emotional voices with a high behavioural relevance for the listener (Ethofer et al., 2007).
Task-dependent effects

If the participants’ attention is explicitly directed towards the emotional auditory stimulus by specific task instructions further cerebral regions are activated. Explicit judgement of the emotional valence conveyed by prosody, for example, was linked to increased responses within the right posterior STC (post-STC) and the bilateral inferior frontal cortex as compared to explicit evaluation of the verbal content of identical stimuli (Ethofer et al., 2006b). These findings are in line with results from another experiment (Wildgruber et al., 2005) using semantically neutral sentences (e.g., “The visitor reserved a room for Thursday”), spoken by actors in different emotional tones (happy, fearful, angry, sad, disgusted). Participants either had to name the emotional category expressed by prosody or they had to perform a phonetic control task (identification of the vowel following the first “a” in the sentence). In accordance with the findings during processing of single words, explicit evaluation of emotional prosody at sentence level was associated with activation of the right post-STC and the inferior frontal cortex.

A different pattern of task-dependent effects was observed, when subjects were asked to voluntarily switch spatial attention between the left and the right ear while performing a gender discrimination task during a dichotic listening experiment. Anger prosody presented at the to-be-attended side yielded increasing activation within the medial frontal cortex (MFC) and the medial occipital cortex as compared to presentation of identical stimuli at the to-be-ignored ear (Grandjean et al., 2005; Sander et al., 2005). The right amygdala and the bilateral mid-STC, in contrast, responded to anger prosody irrespective of the direction of spatial attention.

Besides emotional information, speech prosody conveys information about linguistic meaning (e.g., determining if a sentence is a statement, a question or a command). Experimental findings indicate that the contribution of the right and the left cerebral hemisphere to the extraction of acoustic signals depends upon specific stimulus properties (Wildgruber, Ackermann, Kreifelts, & Ethofer, 2006; Reiterer et al., 2005, 2008). According to the acoustic-lateralization hypothesis slow changes of acoustic signals (e.g., modulations of prosody) are processed within the right hemisphere, whereas the left hemisphere is better suited to process rapid changes of acoustic signals (e.g., differentiation of speech sounds at the level of syllables or phonemes). To further disentangle the impact of functional aspects (emotional vs. linguistic) from basic auditory processing, evaluation of emotional and linguistic prosody was compared using acoustically highly controlled stimuli. To this end, the intonation-contour of a semantically neutral sentence (“the scarf is in the chest”) was systematically manipulated by digital resynthesis. Participants were asked to differentiate pairs of these sentences either with respect to their emotional arousal (“which of the two sentences sounds more excited?”) or their sentence focus (“which of the two sentences is better suited to answer the question: where is the scarf?”). As compared to rest both tasks yielded bilateral frontotemporal activation including rightward lateralization at the level of the post-STC (Wildgruber et al., 2004). Analysis of task-effects, however, revealed activation of bilateral orbitofrontal cortices linked to emotional evaluation.

Model of emotional prosody processing

A connectivity analysis of cerebral activation revealed that the right post-STC is the most likely input region into the network of areas characterised by task-dependent activation (Ethofer et al., 2006b). This finding is in line with the assumption that this area subserves the representation of meaningful prosodic sequences and receives direct input from primary and secondary acoustic regions. Moreover, the connectivity analysis indicated a flow of information along parallel projections from the right post-STC to the bilateral inferior frontal cortex. Taken together these findings indicate multiple successive processing stages, during recognition of emotional prosody following representation within the primary auditory cortex (Figure 1). The first step, extraction of supra-segmental acoustic information, is...
associated with activation of predominantly right hemispheric primary and secondary acoustic regions. The second step, representation of meaningful supra-segmental acoustic sequences, is linked to posterior aspects of the right superior temporal sulcus. The third step, emotional judgment, is linked to the bilateral inferior-frontal cortex (IFC). Within this network, the projection from primary acoustic regions (A1) to the secondary acoustic representation within the mid-STC seems to be predominantly stimulus-driven (bottom-up effect), whereas further projections to the post-STC and the IFC depend upon focusing of attention towards explicit emotional evaluation (top-down effects). It should be mentioned, however, that the results obtained from connectivity analyses of fMRI data do not necessarily imply that there is a direct structural connection between the respective areas. The flow of information from the right post-STC to the left IFC, for example, is presumably mediated by further brain regions such as the left post-STC.

Considering implicit processing of emotional prosody, task-independent activation of the amygdala has been observed (Sander et al., 2005; Ethofer et al., 2009). Additionally, activation within the medial frontal cortex has been demonstrated during implicit processing of emotional vocalisations (Sander et al., 2005; Kreifelts et al., 2009a). These observations are in line with the results of a recently published meta-analysis (Amodio & Frith, 2006) that linked activation of the MFC to mentalizing processes (such as evaluating the intentions of the communicative partners).

Crossmodal integration of emotional signals

Considering bimodal integration of emotional prosody and facial expressions, evaluation of audiovisual emotional signals yielded increased activation within the bilateral post-STC and the right thalamus as compared to either of the unimodal stimulations (Ethofer et al., 2006c; Kreifelts et al., 2007, 2009b). Moreover, enhanced connectivity between the bilateral post-STC and voice-sensitive (mid-STC) as well as face-sensitive (fusiform face area) regions during processing of multimodal signals has been observed, which possibly depicts the mechanism of bimodal binding (Kreifelts et al., 2007). Presumably, the formation of such multimodal associations within the post-STC might also contribute to the understanding of unimodal communicative signals. In these instances missing sensory information might be complemented on the basis of established associations from memory to determine the “meaning” of the signal. During social interaction, however, the emotional connotations of communicative signals are usually not explicitly judged. In fact, highly automatic monitoring of emotional information conveyed by various channels of communication is permanently required. A variety of experimental data indicate different cerebral pathways to be involved in explicit and implicit processing of emotional signals (Figure 2). Limbic brain structures sensitive to emotional information (e.g., amygdala, nucleus accumbens), exhibit a more pronounced response during implicit stimulus processing as compared...
to explicit and cognitively controlled evaluation (Harri et al., 2003; Lange et al., 2003). Based on functional imaging studies it has been postulated that dorsolateral and lateral orbitofrontal areas are involved in the inhibition of limbic activation during explicit emotional judgements (Blair et al., 2007; Mitchell et al., 2007). Increased activity within the medial frontal cortex, in contrast, has been associated with enhanced affective evaluation contributing to selection of emotional significant information in accordance with individual motivational goals and current behavioural demands (Kringelbach, 2005; Phillips et al., 2003). Taken together these findings suggest a predominant role of subcortical limbic regions for implicit emotional processing and a stronger involvement of cortical regions (various frontal areas as well as modality-specific sensory areas) during explicit and cognitively controlled processing of emotional stimuli.

Considering crossmodal integration, facial expressions were rated as being more fearful when presented concomitant with fearless prosody. These changes due to implicitly processed fearful prosody were correlated with enhanced activation of the left amygdala (Etohfer et al., 2006d). In a recent experiment, implicit cross-modal integration was evaluated while subject had to perform a gender discrimination task (Kreifelts et al., 2009b). Bimodal stimulation yielded increasing activation of post-STC, thalamus, amygdala and fusiform gyri. Among these areas, however, solely the right post-STC displayed a positive correlation of the individual hemodynamic integration effect and a measure of trait emotional intelligence as well as voice and face sensitivity. Cumulating evidence, thus, indicates that the post-STC might serve as an essential interface between perceptual integration and social cognition. Future research is required, however, to further delineate the complex pattern of interaction effects within the whole network of brain regions contributing to processing of emotional signals conveyed by various means of communication.

References


