Available online at www.sciencedirect.com

SciVerse ScienceDirect

Physics of Life Reviews 8 (2011) 383–403

PHYSICS of LIFE reviews

www.elsevier.com/locate/plrev

Review

Emotional voices in context: A neurobiological model of multimodal affective information processing

Carolin Brück [a,b,*], Benjamin Kreifelts [a], Dirk Wildgruber [a,b]

[a] *Department of Psychiatry and Psychotherapy, University of Tübingen, Calwerstraße 14, 72076 Tübingen, Germany*
[b] *Werner Reichardt Centre for Integrative Neuroscience, Paul-Ehrlich-Straße 17, 72076 Tübingen, Germany*

Received 8 September 2011; accepted 11 October 2011

Available online 19 October 2011

Communicated by L. Perlovsky

Abstract

Just as eyes are often considered a gateway to the soul, the human voice offers a window through which we gain access to our fellow human beings' minds – their attitudes, intentions and feelings. Whether in talking or singing, crying or laughing, sighing or screaming, the sheer sound of a voice communicates a wealth of information that, in turn, may serve the observant listener as valuable guidepost in social interaction. But how do human beings extract information from the tone of a voice? In an attempt to answer this question, the present article reviews empirical evidence detailing the cerebral processes that underlie our ability to decode *emotional* information from vocal signals. The review will focus primarily on two prominent classes of vocal emotion cues: laughter and speech prosody (i.e. the tone of voice while speaking). Following a brief introduction, behavioral as well as neuroimaging data will be summarized that allows to outline cerebral mechanisms associated with the decoding of emotional voice cues, as well as the influence of various context variables (e.g. co-occurring facial and verbal emotional signals, attention focus, person-specific parameters such as gender and personality) on the respective processes. Building on the presented evidence, a cerebral network model will be introduced that proposes a differential contribution of various cortical and subcortical brain structures to the processing of emotional voice signals both in isolation and in context of accompanying (facial and verbal) emotional cues.

Keywords: Emotion; Communication; Voice; Speech prosody; Nonverbal vocalizations; Laughter; Brain

Contents

Abbreviations: fMRI functional magnetic resonance imaging; STC superior temporal cortex; m-STC middle aspects of the superior temporal cortex; p-STC posterior aspects of the superior temporal; DLPFC dorsolateral prefrontal cortex; FFA fusiform face area; TVA temporal voice area; arMFC anterior rostral medial frontal cortex; OFC orbitofrontal cortex; BG basal ganglia.

[*] Corresponding author at: Department of Psychiatry and Psychotherapy, University of Tübingen, Calwerstraße 14, 72076 Tübingen, Germany. Tel.: +49 7071 2986543; fax: +49 7071 294141.

E-mail address: Carolin.Brueck@med.uni-tuebingen.de (C. Brück).

1571-0645/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.plrev.2011.10.002

Whether we feel happy or sad, enraged or surprised, ashamed or proud, contemptuous of or in love with somebody – our emotions seldom go unnoticed. Similar to lighthouses broadcasting (visual and auditory) warning signs to assist navigation [1], human beings emit a broad range of signals that allow insight into their emotional landscape. Body posture, gestures or facial expressions, for instance, are imbued with hints regarding an expresser's affective state. And once a person raises his or her voice the pictures becomes even more complete. Regardless of whether in talking or singing, sighing, screaming or crying, acoustic cues carried by human voices communicate a wealth of emotional information, and voice-based signals including modulations of the tone of voice while speaking (i.e. speech prosody) or nonverbal vocalizations such as laughter, for instance, offer powerful means to express and decipher emotions.

1. Acoustics of emotions

Imagine someone calling you from afar. In only a matter of seconds, the sheer sound of the caller's voice will probably allow you to discern whether he or she is happy to see you and awaiting you with good news, or whether you should rather turn around and walk away because the caller apparently is very angry with you.

Everyday experience teaches us that specific emotions appear to be associated with distinct vocal patterns and voice characteristics. Our voices, for instance, tend to tremble and quiver when we are nervous, yet they may sound harsh, loud and firm when we are angry.

Driven by the idea that emotion-specific sounds of the human voice should be reflected in a characteristic acoustic signature, research has devoted much attention to delineating acoustic parameters that allow to distinguish different types of emotions [2–4]. Early accounts derived from animal studies further emphasize the idea of a rule-based system that links the motivational state of an animal and the physical structure of vocalizations emitted in close social contact (see motivation-structural rules proposed by Morton [5]). Observations in different species of birds and mammals indicate that "natural selection has resulted in the structural convergence of many animal sounds used in 'hostile' and 'friendly' contexts" [5, p. 855]. Surveys conducted to detail commonalities delineate a general pattern proposing that "birds and mammals use harsh, relatively low frequency sounds when hostile and higher-frequency, more pure tone-like sounds when frightened, appeasing or approaching in a friendly manner" [5, p. 855]. Even though with speech, human beings have adopted more sophisticated means of communication, many aspects of the "primitive" affective signaling system described for animals may in fact still pervade human vocal behavior [6].

However, in the course of evolution human vocal communication appears to have undergone "domestication" [6, p. 236] – reflected, for instance, in the effects of social constraints (e.g. norms or display rules) that influence the way human beings express themselves. In result, vocal characteristics of human affective expressions are shaped by the complex interaction of both physiological processes (e.g. changes in respiration or muscle tone induced by emotional arousal) as well as external determinants monitoring and regulating expressive behaviors (e.g. self-presentation processes) [2,3,6,7] – which in turn might render the task of determining characteristic acoustic patterns of emotions all the more difficult [6].

And indeed, reviews of earlier attempts to define an emotion-specific signature in the human voice reveal an apparent paradox between a listener's ability to decode emotional meaning from vocal cues and research's failure to identify a set of vocal indicators that reliably differentiate a number of discrete emotions [8]. The limited success to define emotion-specific vocal profiles, in turn, fuelled the notion that – instead of signalling discrete affective states – the tone of a person's voice might rather "only" index (nonspecific) physiological arousal [1,9,10] to which various different emotions may be attributed to.

Despite the early discouragement, however, continuous research over the years has produced evidence speaking in favour of pursuing the idea that different affective states might be reflected in specific patterns of acoustic cues. Among the large number of acoustic features that have been studied, primarily four parameters have emerged as prime candidates to subserve the vocal signalling of emotions [2–4]:

(1) voice intensity corresponding to the perceived loudness of a given vocal signal;
(2) tempo and pausing corresponding to the rate of vocalization (e.g. speech rate or laughter rate);
(3) fundamental frequency of vocal fold vibration (F0) defining the perceived pitch of a voice;
(4) energy distribution in the frequency spectrum (i.e. relative amount of energy within predefined frequency bands [3]) affecting voice quality.

Rigorous analysis of acoustic cues measured from samples of emotional speech, helped to define distinctive acoustic profiles for a set of central emotions such as anger, joy, fear, or sadness. The respective findings are summarized in the literature as follows [2,4,11–13]: While *anger* generally has been described to be indexed by increasing voice pitch paired with increases in loudness; *sadness*, for instance, has been characterized to show decreases in voice pitch, speech rate and loudness of a speakers voice. *Fear*, on the other hand, has been suggested to be communicated by increases in voice pitch combined with increases in speech rate, while *joy* has been related to increases in voice pitch, loudness and speech rate.

Similar results have been obtained for non-speech vocalizations: Analogous to the encoding of emotions in speech prosody, different affective states expressed by various nonverbal vocalizations have been associated with distinct acoustic profiles that summarize the complex interplay of acoustic properties related to voice quality, pitch and intensity [14,15]. Obtained findings not only underline the importance of similar acoustic cues used to encode different emotional messages, but also suggest some degree of consensus between vocal patterns associated with the same emotional state in speech prosody and non-speech vocalizations [15,16].

But how do human beings extract emotional meaning from vocal cues?

Attempts to answer this question necessarily involve a closer look at the cerebral mechanisms governing the decoding of vocal expressions of emotions beginning with basic stages of voice perception and advancing to the cerebral processes that allow to infer specific emotions from voice-based acoustic signals such as emotional prosody or nonverbal vocalizations.

2. The neurobiology of voice perception: temporal voice areas of the brain

Comparing the different sound structures that surround us day after day, human voices no doubt take a special position. Human beings probably spend more time analyzing or simply listening to voices than any other sound of their auditory environment [17]. Given the amount of training and experience and the relative ease with which most people decode voice cues, human beings, in a sense, can be considered *voice experts* [18].

At the cerebral level this "expertise" appears to be reflected in a heightened sensitivity to human voices reported for the brain's auditory system: Studies contrasting activation patterns associated with the perception of human vocal sounds (e.g. speech, laughter, singing, cries) to brain responses induced by other natural sound structures, such as animal cries, musical or machine sounds, have consistently indicated voice-sensitive increases in activation for cortical regions located along the middle parts of the right and left superior temporal cortex (e.g. [19–22]). In recent years voice-sensitive responses of the temporal lobe have become a subject of thorough scientific scrutiny, and findings suggesting a unique role of the superior temporal cortex (STC) in human voice perception have been replicated multiple times.

However, while regions of heightened voice-sensitivity have consistently been tied to superior aspects of the temporal lobe, the lateralization as well as the exact localization of voice-sensitive areas within the STC has been described to vary considerably among individuals. Peaks of voice-sensitive activation have been found in anterior, or central as well as more posterior aspects of the STC – with locations along middle parts (m-STC) being the most frequently observed [23]. Considering lateralization, comparisons between cerebral hemispheres suggest a greater sensitivity of the right STC, however, patterns of left-lateralized or bilateral responses to human voices have been reported [20,23].

Beyond anatomical characterizations, research has expanded its focus to include the functional role of *temporal voice areas* in a variety of voice cognition skills (i.e. abilities to extract, evaluate, and categorize nonlinguistic information available in voices [24]). Particularly the ability to decode vocal emotional information, or more precisely the contribution of temporal voice areas to the process of inferring emotions from vocal cues, constitute a field of research that has attracted attention over the past years.

Gathered empirical evidence in sum suggest a prominent role of voice-sensitive cortices in extracting emotions from human vocal sounds: Imaging studies have repeatedly demonstrated enhanced hemodynamic responses of temporal voice areas during the perception of emotional as compared to neutral voices [25–28] irrespective of attention focus [29] or task demands [28].

Emotion-related modulations of activation within voice-sensitive brain regions have further been linked to increases in emotional intensity [28] or behavioral relevance [30], and different types of emotions have been tied to spatially distinct response patterns within the voice-sensitive m-STC [26]. Moreover, strong correlations between hemodynamic responses within the m-STC and acoustic markers of emotions such as mean intensity, mean fundamental frequency or duration [25] additionally strengthen ideas regarding a role of superior temporal structures in the perceptual analysis of emotions expressed by voices.

A closer look at the current literature available on the topic, however, underlines that the contribution of temporal voice areas represents just one piece of the puzzle of how the human brain recognizes and appraises vocal emotional information. In fact, the processing of vocal expressions of emotions (e.g. speech prosody) appears to rely on a network incorporating not only voice-sensitive areas, but also posterior temporal, frontal and subcortical brain structures.

3. The cerebral processing of emotional prosody

It's not what you say, but how you say it – a phrase most people perhaps deem clichéd, yet no other saying so concisely grasps the essence of (emotional) speech communication. Considering the conversations one encounters on a day-to-day basis, the manner in which a person is saying something – the prosody of his or her speech – often appears to be much more informative of the speaker's state of mind than the words he or she uses.

Instead of being a mere by-product of talking, speech prosody in itself provides a rich source of information and powerful means to convey affective meaning aside (from) verbal cues. For millennia, orators have recognized the importance of prosodic indicators to communicate affective subtext that may sway and persuade their audience [31]. And even outside the professional rhetoric arena, prosodic emotional signals appear to lie at the very heart of emotion communication: Inquiries to determine which cues are used to tell how someone is feeling put prosodic indicators in a top position and identify vocal signals as the cues most frequently used and most important in inferring the emotional states of our partners of interaction [32].

Even in situations were no information but vocal inflections are available, prosodic indicators alone allow a rather precise decoding of affective states. Picture, for instance, a scenario, in which you overhear two people talking behind closed doors in a language foreign to you. Although no visual or verbal emotional cues are present to you, you will probably still be able to form an impression about the speakers' current state of mind just by listening to the tone of their voices (see for instance cross cultural studies [33,34]).

Numerous studies in the field of vocal affect perception illustrate that listeners, in general, are rather skilled at inferring emotional meaning from vocal qualities of speech [12]. Summary statistics provided by a recent meta-analysis [12], for instance, indicate varying yet high rates of decoding success for various emotions such as anger (Rosenthal and Rubin's effect size index[1] $pi = 0.93$), sadness ($pi = 0.93$), fear ($pi = 0.88$), joy ($pi = 0.87$) or love ($pi = 0.82$) [12].

[1] Rosenthal and Rubin's effect size index *pi* allows to compare accuracy rates among studies by converting obtained scores to a common standard of measurement where 1.0 corresponds to 100% correct decoding and 0.5 marks chance-level performance. Conversions are justified by the fact that the interpretation of accuracy rates needs to take into account the number of response alternatives presented to a subject. Accuracy levels of 0.50, for instance, indicated chance-level performance (i.e. guessing) when subjects are asked to choose between two response alternatives. Yet a value of 0.5 suggests high levels of correct decoding when results are based on a ten-choice response format (for which decoding accuracies of 0.1 would be expected by chance). *Pi* re-calculates any given accuracy rate to its equivalent had answers been based on a response format providing two different choices.

But what constitutes our ability to comprehend prosodic emotional signals at a cerebral level?

Early neuroanatomical models [35] identify prosody processing as a function dominant to the right cerebral hemisphere whose organization in the brain closely mirrors the left-lateralized cerebral representation of language production and comprehension [36,37]. According to these early models, prosody production was assumed to be mediated by a region within the right dorsolateral prefrontal cortex corresponding in location to Broca's area, whereas prosody comprehension was suggested to rely on right superior temporal structures homologous to Wernicke's area.

However, decades of research since have shifted concepts away from the idea of a single processing center towards models regarding more widespread cerebral networks (including superior temporal and frontal as well as subcortical brain structures) as neurobiological bases of prosody comprehension [38–41]. Beyond the "mere" identification of a prosody network, various imaging studies have contributed to characterizing response patterns of several brain regions implicated in prosody processing, and based on the respective findings, hypotheses regarding the roles of each structure have been formed (reviewed in Ref. [39]).

3.1. The role of the mid superior temporal cortex

Compelling evidence regarding an involvement of the mid superior temporal cortex (m-STC) in the decoding of emotional prosody derives primarily from a variety of neuroimaging studies comparing activation pattern associated with the perception of emotional prosody to brain responses acquired during the processing of speech samples spoken with a neutral tone of voice. Data provided by those studies illustrates that, relative to neutral speech, "hearing" emotional prosody enhances hemodynamic responses within the m-STC [26,28–30], regardless of the type of emotion expressed [25]. Recent attempts to determine the exact localization of activation within the m-STC pinpoint the focus of emotion-sensitive responses to structures previously defined as temporal voice areas as well as adjacent aspects of the primary auditory cortex [42]. Findings, moreover, indicate that the processing of emotional prosody not only increases responses, but also enhances the functional coupling of temporal *emotional voice areas* [42] with more distant brain regions including parietal and frontal as well as subcortical structures such as the basal ganglia and thalamus [42].

Aside a refined anatomical definition of emotion-sensitive activation, increasingly more data has become available that allows to detail response characteristics of mid superior temporal regions implicated in prosody processing. Piecing together the individual findings reported in the literature, modulations of m-STC activation emerge as being stimulus-driven in nature and mediated by basic stimulus qualities that define emotional prosody regardless of processing context. The proposed characterization, first and foremost, build on research results demonstrating increases in m-STC activation during the processing of emotional prosody irrespective of attention focus or cognitive demands [27–29]. Such result, for instance, have been provided by work of Ethofer and colleagues using functional magnetic resonance imaging (fMRI): Aiming to further delineate STC activation elicited by speech prosody processing, Ethofer et al. [28] obtained brain activation patterns of healthy young volunteers while processing digital recordings of single adjectives spoken either in a happy, a neutral, or an angry tone of voice. Following each stimulus presentation, participants were asked to either judge the emotional states conveyed by the tone of voice or classify each stimulus according to its emotional word content (positive, negative or neutral meaning). Data analysis indicated that both during the evaluation of speech prosody as well as during word content judgment, stimuli spoken in an angry or happy tone of voice elicited a stronger response of the m-STC as compared to neutrally spoken stimuli. In other words, irrespective of whether a participant's attention was focused on the decoding of speech prosody or word content, the perception of stimuli spoken in an emotional tone of voice induced stronger activation of mid superior temporal structures compared to activation patterns obtained for neutrally spoken speech samples. In an effort to replicate observed findings, we conducted an fMRI experiment during which we collected brain scans of 24 healthy volunteers while "listening" to adjectives and nouns spoken either with a happy, a neutral or an angry prosody. Analogous to the experimental conditions employed by Ethofer and colleagues [28], participants were asked to perform two different tasks: (a) word content judgment, and (b) prosody judgment. For each task, the same set of 60 stimuli (20 of which conveyed a positive meaning, 20 a negative and 20 a neutral meaning; 20 were spoken in a positive, 20 in a neutral and 20 in an angry tone of voice) was used and task order was balanced across individuals. To evaluate effects of emotional prosody common to both processing conditions, contrasts between brain responses to stimuli spoken with an emotional (i.e. happy and angry) as relative to a neutral tone of voice were computed for each measurement session and subsequently averaged across tasks and participants. Similar to the results described above, computed contrasts

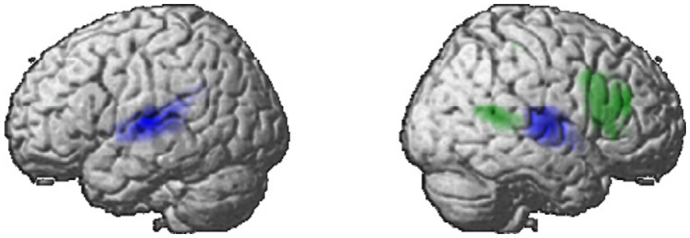

Fig. 1. Stimulus-driven as well as task-related activation associated with the processing of affective prosody. Stimulus-driven activation patterns (marked in blue) were obtained by contrasting speech samples spoken with an emotional (i.e. happy or angry) tone of voice to speech stimuli spoken in a neutral manner (t-contrast: (happy, angry) > neutral, $T \geqslant 3.48$, corrected for multiple comparisons at cluster-level). Please refer to Section 3.1 for further information. Task-related activation patterns (marked in green) were obtained by contrasting brain responses during prosody identification with responses during word content identification (t-contrast: prosody identification > word content identification, $T \geqslant 3.48$, corrected for multiple comparisons at cluster-level). Please refer to Section 3.2 for further information.

revealed a (task-independent) increase of hemodynamic responding within the right and left m-STC driven by the presentation of speech stimuli with emotional prosody (Fig. 1, activations marked in blue).

However, the question arises as to which stimulus-bound features might mediate the observed effects. Evidence to resolve this issue is provided by a recent fMRI experiment seeking to disentangle the effects of various acoustic parameters such as, for instance, mean intensity, mean fundamental frequency or stimulus duration on responses observed within the m-STC [25]: By means of regression analysis, significant linear relationships between hemodynamic responses within the m-STC and acoustic parameters were revealed indicating that modulations of activation within the m-STC are, in fact, driven by the acoustic structure that define emotional prosody at a signal level. However, when analyzed for their individual contribution, none of the chosen acoustic parameters alone sufficed to explain the observed activation patterns. Rather only in modelling the effect of all acoustic parameters combined, increases in m-STC activation during the processing of affective prosody could be predicted sufficiently [25].

In sum, reported findings indicate that:

(1) Increases in m-STC activation during the processing of affective speech prosody are tied to aspects of the temporal lobe implicated in the processing of human voices and sounds in general [17,20,21].
(2) Responses within the m-STC occur to a broad range of vocally expressed emotions irrespective of attention focus or task demands [27,28,30,42].
(3) Activity within the m-STC appears to be mediated by acoustic parameters such as loudness, duration or fundamental frequency.

Taken together the respective findings contribute to the idea that observed m-STC recruitment might reflect stages of perceptual analysis including the extraction and integration of several acoustic parameters into a single emotional percept [25,39] which in turn may provide a groundwork further steps of processing (bound to other nodes of the prosody network) are based upon.

Whereas activation of the m-STC proved to be independent of task demands or attention focus, response patterns of several other structures implicated in prosody processing have been reported to be task-related and linked to task instructions, which required participants to focus on identifying the type of emotion expressed in a given prosodic signal.

Task-dependent response properties have been described for several activation clusters observed within the posterior temporal and frontal cortex as well as limbic regions.

3.2. The role of posterior superior temporal and frontal structures

Aiming to delineate cerebral correlates of prosody processing, several investigations have relied on experimental paradigms that target the respective processes by comparing brain activation related to task conditions requiring the "labelling" of expressed emotional states with experimental conditions that distract attention away from prosodic emotional cues, yet resembled the prosody task with respect to sensory input or basic response mechanisms. To further illustrate the approach, let us revisit the fMRI experiment used to replicate findings on stimulus-driven m-STC activation described in Section 3.1. Instead of contrasting brain responses between different stimulus categories (i.e. emotional prosody vs. neutral prosody), data analysis can easily be adapted to focus on task-related effects. As both conditions share basic parameters such as stimulus material or number of response alternatives, and thus only differ with respect to the process of interest (prosody identification vs. word content judgment), systematic comparisons between both tasks may allow to isolate activation patterns specific to each of the two cognitive operations. To this end brain responses are averaged across all stimulus presentations within each task, and resulting task-related activations are compared by subtracting the respective activation maps from each other. Employed subtraction analysis reveals that in contrast to word content judgement, prosody identification more strongly recruits the right posterior superior temporal cortex (p-STC) as well as the right dorsolateral prefrontal cortex (DLPFC) (Fig. 2, activations marked in green).

Obtained results prove to be in accordance with a number of studies adopting similar task comparisons to explore prosody processing: In fact, across the respective studies one of the most consistent finding appears to be that the explicit evaluation of prosodic information enhances cerebral activity in regions of the right posterior superior temporal lobe as well as the right and left dorsolateral frontal cortex [27,43–47]. However, in addition to activation of the p-STC and DLPFC, lesion data [48,49] as well as imaging evidence [27,45,46] further suggests an involvement of the orbitofrontal cortex (OFC) in prosody identification.

As far as the functional roles of posterior superior temporal and frontal structures are concerned, the following hypotheses are proposed.

Based on similarities between the anatomical localization of dorsolateral frontal structures implicated in prosody decoding and prefrontal constituents of a brain network subserving working memory[2] [51–53] DLPFC activation might be assumed to reflect an engagement of working memory processes relevant to the decoding of affective speech cues [54] and decision making in general.

However, as far as the role of adjacent orbitofrontal structures is concerned, a more emotion-specific contribution has been suggested in the literature: Comparisons between the processing of emotional and linguistic cues carried by speech prosody, for instance, demonstrate that while similarities emerge with respect to the activation of dorsolateral frontal and superior temporal brain structures, recruitment of the OFC remains restricted to processing conditions requiring the evaluation of emotional information [46]. Finding of an emotion-related OFC involvement are further corroborated by imaging data demonstrating OFC activation specific to the evaluation of not only vocal [27,46], but also visual [55] and gustatory [56] stimuli of affective meaning or value, as well as lesion data indicating a selective impairment of emotion identification in patients with OFC damage [48,49]. Building on the lesion and imaging evidence, as well as neuro-anatomical considerations – particularly the OFC's rich connections with sensory cortices and emotion-processing limbic areas [57] – the OFC has been suggested to subserve the evaluation of (visual, vocal, olfactory/gustatory) emotional stimulation possibly by forming associations between perceived emotional signals and episodic memory [40].

Considering the role of the p-STC, however, findings linking posterior superior temporal structures to the integration of visual and vocal emotional information (e.g. face expressions and voice signals, please refer to Section 5.1 for further details) may lead to the proposition that p-STC activation reflects processing stages aiding the identification of emotional information by means of multimodal binding [36]. At first glance, ideas relating multimodal integration to the identification of (unimodal) affective signals such as speech prosody no doubt may tend to baffle or confuse. However, the respective notion draws on the assumptions that, even in situations where no sensory input from other channels of communication is available, additional cues might be retrieved from memory on the basis of established multimodal associations, and through recalling and integrating those retrieved cues with presented prosodic signals additional information becomes available upon which to base emotional inferences. In other word hearing emotional voices during, for instance, a phone conversation might cue memories of matching facial expressions, which one has encountered before in a similar context. Listening to a happy voice might thus, for example, almost automatically evoke mental images of smiling faces, due to the experiences that those particular facial and vocal signals have often been closely linked in the past. As more information becomes available, the identification of conveyed emotional messages might be facilitated and most likely may increase in accuracy. At the cerebral level, this process of retrieving and integrating matching facial cues might not only be mediated by the p-STC itself, but might also rely

[2] In keeping with the definition of Baddeley (1998) "the term *working memory* refers to a system that has evolved for the short-term maintenance and manipulation of information (p. 234)" crucial to many cognitive functions including, for instance, planning and problem solving [50].

on modality-specific visual processing modules, such as the fusiform face area (FFA) [58]. Beyond its implication in perceptual processes, the FFA has also been associated with the mental imagery of faces [59]. Proceeding from this observation, one might assume that "images" triggered by listening to emotional voices are supported by activation of the FFA, and via a complex interplay between this modality-specific processing module and integration sites within the p-STC mental images are combined with perceived voice cues maximizing the information available. However, the presented ideas regarding a contribution of multimodal binding to speech prosody decoding at this point of time remain hypothetical in nature and future research is warranted to further evaluate the claims.

In contrast to increases in activation reported for frontal as well as superior temporal brain areas, task instructions requiring participants to focus attention on the evaluation of prosodic speech cues may also result in a *de*activation of certain brain structures. Particularly, response patterns observed for subcortical limbic regions such as the amygdala provide empirical foundation to claims of task-related decreases in hemodynamic responding.

3.3. The role of limbic structures

Considering amygdala activation associated with prosody decoding, seemingly contradicting pieces of empirical evidence have been published in the literature. Reports of an increased [60] as well as decreased [61] or no "critical" [62] involvement during the processing of vocally expressed emotions leave the role of the amygdala in speech prosody processing an issue of debate. However, across studies a pattern begins to emerge: Enhanced activation of the amygdala often appears to be associated with task conditions (e.g. speaker identification) requiring a more pre-attentive or implicit[3] processing of prosodic signals [27,60,63,64], whereas, in contrast, studies focusing attention on the explicit[4] evaluation of vocally expressed emotions rarely report increases in amygdala activation, but rather indicate strong frontal responses [43,46,47]. Based on these observations one might conclude that: While processing prosodic signals outside of attention focus may result in an increased responding of limbic areas, explicit attentional processing might, in comparison, lead to a deactivation of limbic structures. Latter assumptions are further corroborated by findings suggesting frontal brain structures to inhibit limbic responses during the performance of demanding cognitive tasks [65,66].

Beyond a characterization of task-related effects, reviewed research findings outline the idea of two distinct modes of speech prosody processing, each differentially represented in the human brain:

(1) *explicit* or cognitive controlled processing associated with focusing attention on the evaluation of prosodic speech elements.
(2) *implicit* processing, a rather "automatic" or incidental way of processing, during which speech prosody is decoded without deliberately directing attention towards prosodic signals and many times even without conscious awareness.

As far as the cerebral correlates of both modes of processing are concerned, experimental data, in sum, indicates a role of dorsolateral and orbitofrontal regions as well as posterior superior temporal structures during the explicit decoding of emotional signals, whereas implicit modes of processing have been suggested to engage limbic pathways [67–69] including the amygdala [60,63,64,70], as well as aspects of the anterior rostral mediofrontal cortex (arMFC) [70,71] assigned a fundamental role in "mentalizing" [72] (i.e. the ability by which knowledge about the desires and beliefs of others is acquired [73]).

Ideas of a "limbic" processing path on the one hand, and a "cortical" pathway on the other hand present to be in accordance with a classic model of emotional information processing proposed by Joseph LeDoux [74]: LeDoux argues that following basic stages of sensory analysis within the thalamus, emotion processing relies on two distinct neural circuits: (1) the "low road" – a pathway connecting the thalamus directly with the amygdala and (2) the "high road" relaying information from the thalamus to the cortex and on to the amygdala. The low road is described to represent a "safety system" bypassing conscious awareness that, in a quick and efficient way, triggers emotional

[3] The term "implicit" is used to describe conditions during which the processing of prosodic emotional signals is task-irrelevant and occurs rather incidentally or unintentionally.

[4] "Explicit" refers to processing conditions during which the evaluation of prosodic emotional signals is task-relevant and attention is devoted to interpreting emotions expressed by speech prosody.

responses which alert and prepare the body to act. The second loop – the so-called high road – in comparison relates to slower, yet conscious and much more elaborate way of processing that create a more precise representation of the environment and a more thorough appraisal of the situation at hand, and via projections to the amygdala this pathway allows to control and fine-tune emotional responses triggered by the "low road" to fit the needs of a given environment.

Applying LeDoux's ideas to prosody processing, the following hypotheses can be deduced: While limbic activation tied to implicit modes of prosody processing may reflect the automatic induction of emotional responses, cortical activation patterns observed during explicit processing condition by contrast may be linked to cognitive control and appraisal processes which help to regulate limbic responses and above all allow a thorough evaluation of the prosodic cues send by our partners of interaction.

3.4. The role of the basal ganglia

Aside limbic regions, a second group of subcortical brain structures – the basal ganglia (BG) – have frequently been implicated in prosody decoding. Particularly, research in patients diagnosed with Parkinson's disease (PD), a condition marked by the degeneration of the BG, has drawn attention to BG involvement in speech prosody processing: Compared to healthy individuals, PD patients often exhibit severe difficulties in deciphering emotional meaning from prosodic speech cues: Faced with identifying different emotional states from speech samples, PD patients are prone to error and tend to misjudge vocal emotional signals with respect to the type of emotion expressed or more general properties such as emotional valence[5] [75–79]. Evidence gathered from PD patient samples is further supported by reports of impaired prosody decoding following traumatic lesions of the BG [80–83], as well as imaging studies indicating BG activation during the processing of prosodic emotional cues [63,84].

However, despite the growing body of research, the exact role of the BG in prosody processing remains elusive: While most findings speak in favor of an involvement in cognitive–evaluative processes associated with prosody identification (e.g. [75–79,82,83]), evidence has been presented that suggests a role of the BG in the processing of acoustic markers of vocally expressed emotions such as speech timing [85]. Among the many issues that merit further clarification, one of the most fundamental questions appears to be whether observed BG involvement reflects a contribution specific to speech prosody processing or not. In fact, one may argue that the functional relationship between speech prosody decoding and the BG can be explained by the BGs crucial involvement in basic executive functions such as focussing and sustaining attention or planning and regulating behavior, as well as working memory [86]. Such theoretical considerations are supported, for instance, by empirical evidence indicating that speech decoding deficits might not be restricted to emotional prosodic signals per se but may rather apply to a broader range of vocal information including linguistic prosody [87], and are at least in part attributable to disturbances of working memory [85]. In sum, future research is encouraged to further specify the contribution of the BG in the process of speech prosody decoding.

3.5. Section summary: a cerebral network model of prosody processing

Taken together research results presented in this section define prosody comprehension as a complex function tied to several cortical and subcortical brain structures. Research has established a contribution of brain structures including the p-STC and m-STC, the DLPFC and OFC as well as limbic regions such as the amygdala or aspects, of the arMFC. Each of these brain structures, in turn, has been suggested to be associated with distinct aspects of prosody decoding from basic stages of acoustic analysis to higher-order evaluative processes.

The localization of cerebral correlates, in addition, is complemented by empirical evidence that allows to further detail the complex interplay of brain structures implicated in prosody decoding. Investigations of interactions among different nodes of the prosody network point to a strong coupling of frontal and temporal activations, and suggest a flow of information from the right STC to the right and left DLPFC during the processing of emotional speech prosody [43].

In sum, current research findings propose the idea that prosody comprehension is mediated by a sequential multistep process unfolding from basics stages of acoustic voice analysis (bound to temporal brain areas) and proceeding to

[5] The term valence in this context is used to refer to the "pleasantness" of an emotional state. Judgments of valence often require to evaluate whether a given emotional signal represents a more positive or negative state of mind.

higher-level stages of categorization and recognition (associated with frontal aspects of the brain). Following the processing of auditory information within the ear, brainstem, thalamus and primary acoustic cortex (A1), three successive steps of prosody decoding can be identified:

Step 1: extraction of acoustic features of prosodic cues;
Step 2: identification of vocally expressed emotion by means of multimodal integration;
Step 3: explicit evaluation and cognitive elaboration of vocally expressed emotions.

Each of these steps, in turn, appears to be differentially represented in the human brain: Whereas the extraction of acoustic features has been linked to voice-sensitive structures of the m-STC, more posterior aspects of the right STC have been recognized for their contribution to the identification and integration of emotional signals into a common percept (please refer to Sections 3.1 and 3.2). On the other hand, sub-processes concerned with the evaluation and cognitive elaboration of vocally expressed emotions have been linked to frontal structures (DLPFC/OFC) of both cerebral hemispheres. Within the prosody network, information transfer from primary acoustic regions (A1) to the mid-STC has been characterized to be primarily stimulus driven in nature, whereas projection to the p-STC and DLPFC have been described to depend upon focusing attention on the explicit evaluation of expressed emotions.

The sequential nature of the proposed processing steps is further corroborated by electrophysiological studies that allow to discern the time course of various aspects of prosody decoding. Recordings of event-related potentials (ERP), for example, link the acoustic analysis of a given speech cue to changes in brain activation occurring within the first one hundred milliseconds following stimulus onset, while processes related to the evaluation of emotional meaning appear to be reflected in variations of brain responses with higher latencies: Research findings [41], for instance, indicate that modulations of acoustic properties such as frequency and intensity affect ERP components that peak as early as approximately 60–80 ms [88] following the onset of an auditory event, thus underlining the assertions that the analysis of the auditory input signal occur rather early in the process of speech prosody decoding. However, experimental conditions requiring participants to identify emotions based on prosodic markers of a speaker voice have been shown to modulate ERP signals around 360 ms post-stimulus onset [89], suggesting processing steps concerned with the explicit evaluation of emotional information to succeed stages of acoustic analysis in time.

Aside the explicit decoding tied to the three processing steps detailed above, a second implicit mode of prosody processing has been assumed. In contrast to the explicit evaluation, implicit processing is believed to occur without devoting attention to the interpretation of prosodic signals. Considering cerebral correlates, studies suggest a network including the amygdala and arMFC to subserve the implicit analysis of prosodic speech cues (please refer to Section 3.3). Beyond the identification of two different pathways of processing, recent neuroimaging studies indicate a complex interaction between structures of both processing routes: Findings support the idea that frontal cortical areas implicated in the explicit analysis of prosodic cues might also contribute to the inhibition of limbic activation when individuals actively attend to emotional signals. Suppression of limbic activation, on the other hand, has been assumed to reflect a recruitment of emotion regulation processes which may attenuate the automatic induction of emotional reactions associated with limbic activation to avoid emotional interference in goal-directed behaviour [66].

Questions, however, remain as to how complete the proposed model of prosody processing is. While, at this point of time, the presented account may provide a framework to understand the contribution of fronto-temporal and subcortical brain structures most consistently implicated in speech prosody decoding, reports of a contribution of other brain structures such as, for instance, the BG outline the need for further specification and extension of current knowledge.

4. The cerebral processing of nonverbal vocalizations: the example of laughter

Aside speech, human vocal emotional communication relies on a myriad of nonverbal vocalizations. We, for instance, scream or sob, groan, growl or sigh to express how we feel. And although, due to the resemblance with animal calls [90], one might be inclined to think of nonverbal vocalizations as primitive means of communication, the information carried by these signals often appears to be just as valuable as knowledge gathered from speech or speech-related vocal phenomena such as affective prosody.

When considering the different types of nonverbal vocalizations frequently used in social encounters, laughter no doubt will be among the first that springs to mind. From shy giggles to roaring hearty guffaws, laughter pervades human interaction, shapes and defines social relationships, unites and divides people [91]. A laugh invites a listener to join in and share the joy of its sender. Yet at the same time it may serve as a weapon of social scorn, ridicule and humiliation [91,92].

Although commonly regarded as a sign of happiness and joy, laughter accompanies a much more diverse range of emotional states. Picture, for instance, the nervous giggles of public speaker who has lost his/her train of thought; or the sorrowful almost weeping laughter of a mourner caught up in memories of his lost loved one. In each instance, the listener may hear laughter – a short burst of staccato-like syllabic vocalizations (e.g. ha–ha–ha, ho–ho–ho, he–he–he) [93], yet the laughter associated with the different scenarios not only carries different meaning it also sounds very distinctively. Similar to speech prosody each emotional state appears to be associated with a characteristic sound of laughter [14,94] (please refer to Section 1). Thus, whether a person, for instance, feels joy or contempt towards his or her communication partner, might be discernable solely on the basis of the acoustic signals that constitute human laughter. And indeed, even if no information but the sheer sound of laughter is available, human beings are still able to infer and classify the emotional states of their partners of interaction with remarkable accuracy, as for instance Szameitat and colleagues [95] illustrate: In order to evaluate the role of laughter as a communicative signal, Szameitat and co-workers designed an experiment that sought to investigate whether healthy young adults would be able to determine a person's current emotional state based on the sound of his or her laughter alone. In preparation of this experiment, the authors recorded actors' portrayals of different types of "emotional laughter". To aid the acting performance, actors were instructed to imagine or recall one of four mental states or situations: (1) being tickled or (2) being happy as well as (3) taunting somebody or (4) enjoying the misfortune of another person. As soon as they "felt the mood" of the given situation, actors were encouraged to produce laughter sounds typical of their respective state of mind and the emotions they felt. Each uttered laugh was digitally recorded, which resulted in a set of laughter sequences comprising multiple examples of four basic types of laughter: (1) tickling and (2) taunting or derisive laughter, as well as laughter conveying (3) joy, and (4) schadenfreude. The authors then asked a group of 72 volunteers to classify the recorded laughter sequences according to the emotions conveyed by each sound bit. Obtained behavioral results indicated that despite the fact that no further information but the sheer sound of laughter was provided, listeners nevertheless were able to infer expressed emotional states with accuracies well above chance level for each type of laughter (chance level: 25% correct identifications, joy: 44%, tickling: 45%, taunting: 50%, schadenfreude: 37%).

Although only few studies to date have sought to investigate the neurobiological basis of laughter perception, gathered evidence nonetheless suggests cerebral structures similar to those implicated in prosody decoding: An involvement of temporal brain structures, particularly the superior temporal cortex, as well as frontal and limbic regions has been described in the literature [71,93,96–98]. Moreover, different types of laughter have been associated with distinct cerebral activation patterns: While joyous and taunting laughter (i.e. emotional laughter), for instance, have been shown to elicit particularly strong responses of the arMFC; laughter induced by tickling, in comparison, has been related to a more pronounced recruitment of voices-sensitive structures of the right m-STC [71]. Differences in m-STC activation have been assumed to reflect variations in acoustic complexity among different types of emotional laughter, while differential activation of the arMFC has been assumed to relate to differences in social meaning [71]. Unlike various types of "emotional" laughter that may carry different social messages (e.g. laughter meant to taunting or rejecting somebody, joyous laughter inviting bystanders to join in), laughter induced by tickling can be considered a more unequivocal cue related to playful interactions that promote social bonding [71]. Considering the aforementioned differences in social meaning, one may assume that the processing of emotional laughter may posit a greater challenge on social cognition, and thus more strongly activate cerebral structures implicated in social-cognitive processing such as the arMFC [72].

However, given the limited data available on the topic, current insights into the cerebral mechanisms that govern the decoding of laughter signals can only be considered fragmentary at best. In consequence, future research needs to be devoted to specifying current knowledge: For instance, although at first glance many commonalities emerge between the perception of emotional speech prosody and laughter sounds, direct comparisons between the processing of both types of vocal emotional signals may help to further advance our understanding.

5. The integration of vocal, verbal and facial emotional information

Considering social interaction in everyday life, we rarely encounter vocal emotional cues in isolation. Rather, emotional voice signals are often accompanied by verbal information or visual cues (e.g. facial expression) providing further evidence regarding a person's current state of mind.

Instead of focusing on cues derived from a single channel of communication, successful social interaction appears to rely on the combination or integration of verbal, vocal and visual emotional signals available in a given social situation: Combining multiple sources of information may, on the one hand, facilitate emotion judgments – as evidenced, for instance, by behavioral studies [99–102] indicating shortened response latencies and higher classification accuracy for stimuli conveying matching visual and vocal signals. However, on the other hand, information gathered from one channel of communication may also modulate or alter the interpretation of emotional cues conveyed by another: While, for instance, the sentence "*Oh, I am so happy for you*!" might generally be considered a signal of positive feelings towards the addressee, speaking it in an angry tone of voice will probably evoke a very different "reading" of the statement.

In recent years, an increasing number of studies have aimed at defining the cerebral mechanisms that contribute to combining verbal, vocal and visual affective information. Particularly, the integration of vocal and facial emotional signals has attracted much attention.

5.1. Integrating nonverbal vocal and facial emotional cues

Groundbreaking insights into how the brain processes audiovisual information, or more generally multisensory stimulation have, particularly, been gathered from anatomical and electrophysiological studies both in humans and animals (reviewed in [103–105]). Building on the results obtained in these studies, numerous brain regions might be selected as potential "sites" of audiovisual integration: Sensory-specific brain areas such as, for instance, the auditory cortex [106–108] as well as "convergence zones" [109] – brain areas that receive afferent input from several senses – such as the superior temporal [110,111] and orbitofrontal cortex [110,112], the insula [113], the superior colliculi [114], the claustrum [115], thalamus [116] or amygdala [117] may all be considered prime candidates.

Knowledge derived from electrophysiological studies is further complemented by a growing body of fMRI research. And even though approaches taken to explore audiovisual integration tend to differ considerably among studies, a pattern begins to emerge that suggest a crucial role of the p-STC in the integration of facial and vocal emotional cues. Compelling evidence for an involvement of the p-STC is presented, for instance, in a series of fMRI experiments conducted by Kreifelts and collaborators [99,118,119]. Aiming to delineate audiovisual integration sites in the human brain, Kreifelts et al. chose to contrast brain activation to audiovisual emotional stimulation with brain responses evoked by unimodal (visual or auditory) stimuli. Participants were asked to classify a range of audiovisual (video clips), visual (mute videos) as well as auditory (sound clips) stimuli according to the emotions conveyed by each stimulus. Each video or sound clip depicted men and women expressing different emotional states by means of verbal, vocal and facial signals, and each participant was instructed to label the type of emotion based on the nonverbal cues (facial expressions, tone of voice) presented. Audiovisual and unimodal processing conditions were contrasted with respect to brain activation and behavioral data (i.e. reaction times, accuracy of decoding) allowing to detail effects of audiovisual integration both at a behavioral and cerebral level. Compared to unimodal processing conditions, the audiovisual presentation of nonverbal emotional signals was associated with a significant perceptual gain evidenced by markedly higher accuracy rates obtained during audiovisual processing conditions. At the cerebral level, observed increases in decoding accuracy proved to be associated with increasing activation of the right and left p-STC during the processing of audiovisual nonverbal cues – suggesting activations of the p-STC as cerebral correlates of audiovisual integration [99].

Considerations of connectivity further support the idea of the p-STC as site of audiovisual integration. Not only do projections from both primary auditory and visual cortices converge within the p-STC, but analyses employed to estimate the functional coupling of regions implicated in the processing of nonverbal emotional signals also reveal a functional connection between the bilateral p-STC and temporal voice areas as well as face-sensitive aspects of the fusiform gyrus during the processing of audiovisual nonverbal signals [99]. Building on the observed enhanced functional coupling between unimodal associative cortices and the p-STC during the processing of audiovisual emotional stimulation, the idea has been proposed that audiovisual integration can be envisioned as process during which information from the spatially distinct modality-specific processing sites (e.g. fusiform face area [58], temporal voice areas [20]) is transmitted to integration areas within the p-STC where the information is bound into a single percept [99,105]. In sum, current empirical evidence suggest that the integration of vocal emotional information with co-occurring facial signals might rely on a common "supramodal" processing steps tied to brain structures implicated in audiovisual integration in general (e.g. p-STC) as well as modality-specific steps of processing associated with the recruitment of "unimodal" cortices [104].

In uniting the outlined idea with current data regarding the processing of emotional facial expressions as well as emotional vocal cues, a working model of affective face–voice integration can be conceived. Building on similarities between the processing of emotions in both sensory modalities such as

(a) the reliance on modality-specific "processing modules" (i.e. fusiform face area [58], temporal voice area [20]) contributing to basic stages of perceptual analysis;
(b) an involvement of dorsolateral and orbitofrontal brain structures during the explicit evaluation of both visually and vocally expressed emotions (evidence reviewed in [39,120]);
(c) two routs of processing: explicit "cortical" as well as an implicit "limbic" processing (evidence reviewed in [39, 120]) reported for both emotional face processing as well as emotional voice processing

one might proceed to propose three (general) processing steps associated with the perception of audiovisual nonverbal emotional signals [39]:

(1) extraction of visual and vocal communicative signals within the respective modality specific primary cortices and specialized processing modules;
(2) integration of auditory and visual information into a single percept within the p-STC;
(3) cognitive elaboration and explicit evaluation of emotional information related to the activation of dorsolateral and orbitofrontal brain structures.

Moreover, in analogy to concepts reviewed for speech prosody decoding as well as in accordance with the published literature on the (cerebral) processing of emotional facial expression (reviewed elsewhere e.g. [120]), the working model furthermore assumes that the processing of audiovisual emotional information can occur in two different ways – each tied to different neural circuits and functional meaning: Audiovisual emotional signals may, on the one hand, be decoded in an explicit, cognitive controlled way (represented by the three steps described above). On the other hand, however, such nonverbal emotional cues may be processed in a rather implicit or un-intentional manner implemented mainly via limbic pathways, particularly the amygdala, as well as "mentalizing regions" of the arMFC [72].

5.2. Integrating verbal and nonverbal vocal emotional cues

In human acoustic communication, nonverbal vocal emotional signals (e.g. prosodic signals) are closely intertwined with verbal affective cues (i.e. information defined by the words we use). Both verbal and nonverbal affective information not only co-exist in spoken language, but rather they interact in a complex way to shape affective judgments. Nonverbal vocal cues may strengthen or weaken the verbal messages presented to us. Vice versa, verbal affective signals may modulate our impressions of a speaker's nonverbal vocal cues.

Although research pertaining to the integration of verbal and nonverbal vocal affective signals looks back on a long tradition,[6] insights into the cerebral processes that underlie the ability to unite *what* is said with *how* it is said remain rather limited. Tentative clues, however, might be derived from a few imaging studies focusing on how the brain monitors and resolves conflict between different sources of emotional information (e.g. [44,121,122]).

Inconsistencies between verbal emotional signals and accompanying prosodic cues pose a "decoding problem" human beings are frequently faced with during day-to-day social interaction. Our understanding of rhetoric devices such irony, for instance, or even our ability to detect lies may crucially depend upon noticing and weighing discrepancies between the verbal and nonverbal vocal message presented to us.

[6] See for instance studies conducted by Mehrabian and colleagues in the 1960s and research following up on their findings.

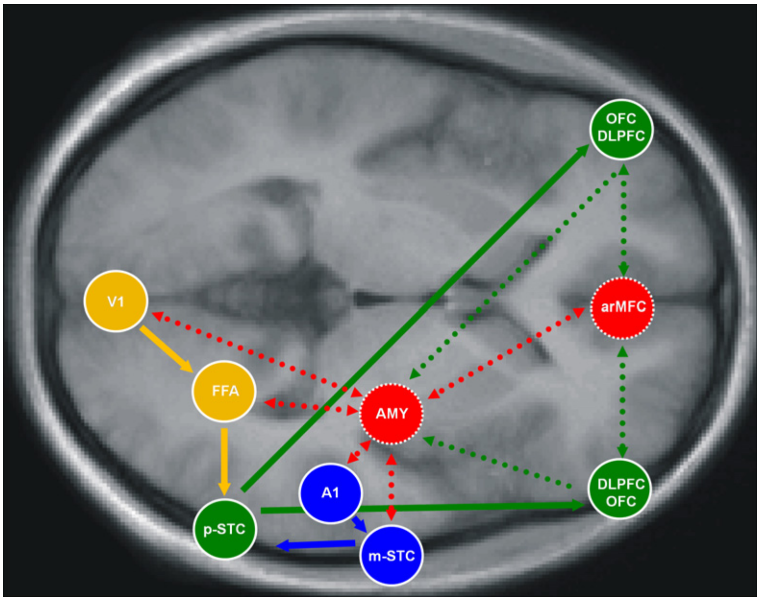

Fig. 2. Cerebral network model of the integration of visual and vocal affective information. The model proposes three steps of processing: (1) The extraction of basic acoustic and visual features tied to modality – specific primary cortices and specialized processing modules. Visual features are processed along a pathway projecting from the primary visual cortex (V1) to face processing modules of the fusiform gyrus (FFA) (see processing path marked in yellow). Acoustic features are analyzed within the primary auditory cortex (A1) and relayed to voice-processing areas of the mid superior temporal cortex (m-STC) (see processing path marked in blue). (2) The integration of auditory and visual information: Following its processing within sensory-specific brain areas, visual and auditory information is transferred to integration sites of the p-STC were it is bound into a single emotional percept and made available to evaluation (see processing path marked in green). (3) Cognitive elaboration and explicit evaluation of emotional information associated with the activation of the dorsolateral frontal and orbitofrontal cortex (DLPFC/OFC). Aside the explicit processing of audiovisual emotional signals (summarized by the three steps above), the model assume a second, implicit way of processing (marked in red) bound to limbic structures such as the amygdale (AMY), as well as "mentalizing regions" of the anterior rostral medial frontal cortex (arMFC). Hypothetical interactions between cerebral structures are marked with bidirectional dotted arrows. Moreover, inhibitory connections (dotted unidirectional arrows) are assumed between the frontal cortex (DLPFC/OFC) and limbic structures. Please note that depicted connections do not necessarily imply direct neuronal connections between different regions, rather the flow of information might be mediated through additional neuronal structures.

In an effort to understand how the brain deciphers speech stimuli combining both sources of information, comparisons between cerebral responses to congruous and incongruous speech samples[7] may provide first evidence as to which brain structures contribute to the decoding process. Research results, for instance, indicate that in contrast to the processing of "pure" prosody (i.e. prosodic speech cues "robbed" of semantic content by filtering speech frequencies above 350 Hz), decoding of congruous stimulus combinations more strongly activates the left inferior frontal as well as the superior and middle temporal gyrus [121]. Processing incongruous stimuli combinations, on the other hand, has been associated with activation of similar regions of the frontal and temporal cortex as well as structures, such as the anterior cingulate cortex, frequently linked to conflict resolution in a broader (non-emotional) context [121,122]. Moreover emotion-specific differences in brain-responses to incongruous stimulation have been described [122]: While the processing of conflict between happy prosody and negative word content appears to more strongly recruit left superior temporal and frontal cortical areas, inconsistencies between angry prosody and positive verbal signals have been shown to modulate activation of subcortical brain structures such as the thalamus and basal ganglia (e.g. caudate nucleus).

Judging from the empirical evidence available one might conclude that, compared to the processing of nonverbal vocal signals in isolation, the presence of verbal semantic cues (whether they match vocal signals or not) results in an engagement of additional cerebral resources. In some respect observed activation patterns might reflect conflict

[7] Congruous refers to speech stimuli presenting matching emotional cues at a verbal and vocal level (e.g. positive verbal messages paired with a positive (happy) tone of voice). The term incongruous, on the other hand, refers to speech stimuli presenting contradicting cues (e.g. positive verbal message paired with a negative (angry) tone of voice).

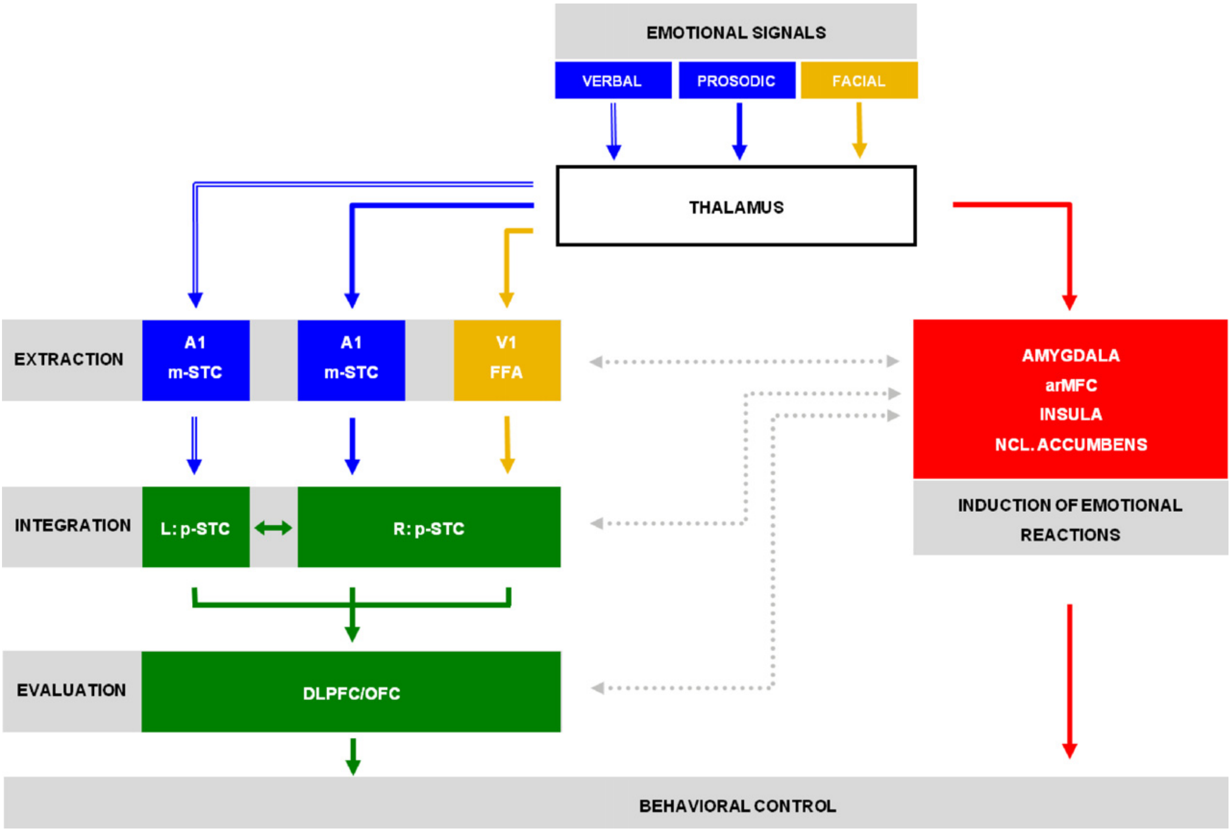

Fig. 3. Extension of the working model of emotional face–voice integration (see Section 5.1) to include the processing of verbal affective cues: Visual, verbal and nonverbal vocal affective information is extracted within modality-specific primary and secondary cortices (visual information = processing path marked in yellow; vocal (verbal and nonverbal) information = processing path marked in blue). Following basic stages of perceptual analysis, information is transferred to integration sites within the right (R) and left (L) posterior superior temporal (p-STC) were it is bound into a single emotional percept. Subsequent to audiovisual integration, emotional information is projected to structures of the dorsolateral frontal and orbitofrontal cortex (DLPFC/OFC) related to the explicit evaluation and cognitive elaboration of presented emotional signals. Aside the explicit evaluation, the model assumes a second implicit mode of processing (marked in red) implemented mainly via limbic structures. Hypothetical interactions between cerebral structures are marked using dotted bidirectional arrows.

monitoring processes implicated in resolving inconsistencies in a wide range of (emotional and non-emotional) tasks [114]. However, emotion-specific effects have also been suggested [121,122].

Moreover, in analogy to the working model of affective face–voice integration proposed in Section 5.1, one may assume that the integration of verbal and vocal affective signals draws on modality – or "channel-specific" mechanisms as well as processing steps implicated in the integration and evaluation of emotional information across modalities. Comparisons between the both channels of communication may allow to further specify the respective hypothesis: As both verbal and nonverbal vocal emotional signals are defined by the acoustic characteristics of a speech sample, similarities between the processing of both types of cues might emerge with respect to early stages of perceptual analysis. Building on evidence presented for the analysis of human vocal sounds (please refer to Sections 2 and 3.1), one may conclude that the extraction of verbal and vocal information relies on the primary acoustic cortex as well as voice-sensitive structures of the m-STC. Considering "channel-specific" aspects of processing, imaging studies suggest distinct networks for the decoding of verbal and nonverbal vocal affective information: While the explicit evaluation of prosodic signals has been related to activation of the right p-STC as well as bilateral frontal brain regions (see Section 3), the decoding of verbal emotional speech cues (irrespective of speech prosody) has been associated with left posterior superior temporal and dorsolateral frontal structures involved in retrieving semantic knowledge, medial frontal structures implicated in mentalizing, as well as subcortical regions including the BG, thalamus and amygdala [123–125]. Differences between the two channels of communication have also been described with respect to activation lateralization: Findings indicate a more pronounced activation of the left hemispheric structures while processing verbal cues, whereas the decoding of prosodic signals appears to more strongly recruit the right cerebral hemisphere [126].

Considering processing steps implicated in the integration of emotional signals, a prominent role of the p-STC has been suggested: Studies concerned with the integration of auditory and visual information consistently identify the bilateral p-STC as a site of audiovisual binding (please refer to Section 5.1). Beyond audiovisual processing, reports implicate the left p-STC in the integration of semantic and syntactic information (see for instance evidence reviewed in [123]). Moreover, current findings suggest that fiber connections between the bilateral p-STC may mediate "interplay" between the right lateralized pathways processing nonverbal information and left lateralized cerebral structures processing verbal information [127]. The respective assumption draws on patient data indicating disturbances of the integration of nonverbal and verbal syntactic information after lesions of posterior aspect of the corpus callosum connecting temporal areas of both cerebral hemispheres [127].

Joining the reviewed empirical findings with theoretical frameworks developed for the processing of vocal emotional cues and face–voice integration (see Sections 3.5, 5.1), the following hypotheses can be deduced (Fig. 3): Building on the acoustic analysis of speech signals within the primary auditory cortex and voice-sensitive m-STC, verbal and nonverbal vocal information is relayed to "integration sites" of the p-STC, bound into a single emotional percept and projected to structures of the DLPFC/OFC associated with the explicit evaluation of emotional meaning. Considering empirical evidence presented on "channel-specific" differences between the decoding of verbal and nonverbal cues, one may assume that the processing of verbal emotional signals predominantly relies on structures of the left cerebral hemisphere (particularly the left p-STC and left DLPFC), while the decoding of vocal emotional signals more strongly engages the right cerebral hemisphere.

6. Similar but different: interindividual variations in the processing of emotional voice cues

Although, in general, common mechanisms of information processing are assumed to apply to each and every member of the human race, the ways in which human beings "process" even similar objects or conditions nonetheless often tend to differ tremendously among individuals. Particularly with respect to emotion processing idiosyncratic discrepancies in how we perceive and respond to the emotional stimuli surrounding us appear to be "the rule rather than the exception" [128]. Most of us, for instance, will probably have encountered situations where the same emotional event (e.g. a sad movie) evokes completely different, maybe even opposing, responses in a group of people (e.g. some break into tears while others remains emotionally untouched) [128]. Such behavioral discrepancies for one may be rooted in disparate life experience or transient mood states that shape human behavior. Yet they may also be linked to more stable individual characteristics such as age, biological sex or personality.

While research in the domain of human affective neuroscience long has been concerned with unraveling common brain mechanisms or regularities in cerebral processes, recent years have brought increasing interest in how differences among individuals might influence brain mechanisms associated with the processing of emotions. And, although still a rather young field of research, gathered findings consistently suggests interindividual differences to translate into substantial variance in brain responses. Several determinants of individual variation such as differences in dispositional affect or personality, hormonal or genotype differences have been shown to modulate the activation of various components of the "emotional brain" [129] including, for instance, the prefrontal cortex, the amygdala and other limbic structures (see evidence reviewed in [128]).

Considering the processing of emotional vocal cues, however, only few studies to date have focused on interindividual variations. Nevertheless, even the relatively sparse data available on the topic still provides valuable additional insights that may complement the wealth of knowledge gained from group-averaged analyses. Current findings, for instance, indicate sex differences in activation amplitudes of frontal brain structures [130] implicated in the processing of emotional speech; as well as sex-related variability in the time-course of brain activity changes attributable to the presentation of prosodic cues [131]. Moreover, first evidence demonstrates a strong influence of personality on brain activation related to the processing of vocal emotional cues [132]. Correlations between measures of personality and activation patterns reveal brain responses to vary as a function of individual differences in neuroticism – a personality trait related to emotional (in)stability [133]. Neuroticism-dependent modulations have been described for activation clusters observed within the arMFC, postcentral cortex as well as the amygdala and anterior cingulate cortex [132], and the respective findings, in turn, have been interpreted to reflect processing biases or differences in task strategies or emotion control associated with different personality profiles [132]. Irrespective of their functional meaning, however, observed sex-related or personality-dependent variance again underline the modulatory role of individual characteristics – rendering brain activation patterns among individuals, in a sense, similar but different.

Proceeding from the initial observation of individual variance in brain activation, a myriad of research questions ensues, e.g.: Can brain activation patterns be used to classify normal and dysfunctional voice perception? Can we distinguish persons skilled in the decoding of vocally expressed emotions from those individuals with difficulties in interpreting such social signals based on typical patterns of brain activation? In other words, pursuing an individual-difference approach in the study of emotional voice processing not only holds the potential to further advance current models, but it might also guide the way to a better understanding of individual disturbances of emotional voice processing associated, for instance, with a variety of psychiatric diseases such as schizophrenia [134], depression [135–138] or autism [139–142]. Important first steps have been taken to elucidate the cerebral bases of dysfunctional voice decoding: Both altered lateralization of temporal lobe responses [143,144] as well as alterations in STC activation [145,146] have been described for patient samples exhibiting difficulties in voice decoding–outlining the assumption that impairments might originate in early stages of voice analysis tied to structures of the temporal lobe. Pending further study of the issue, hypotheses derived from this line of research, not only may allow to understand the nature of observed disturbance, but, in consequence, may also contribute to the development and improvement of therapeutic interventions aimed to ameliorate voice decoding impairments and the dire consequences on social functioning associated therewith. As for deficits in the perception of emotional facial expressions, therapeutic interventions in the form of recognition trainings have been shown to significantly improve recognition abilities [147,148], and the specific training effects have been related to changes in the activation of occipital, parietal and frontal structures possibly reflecting a "more efficient use of attentional, perceptual, or cognitive strategies" [149]. The efficacy of facial affect recognition trainings all the more encourages to further extend therapeutic efforts to other domains of social perception including emotional voice decoding in hopes to achieve similar effects in overcoming disease-related voice decoding difficulties.

7. Summary

Human beings not only feel emotions – they express them in many ways. Through gestures, facial expressions, and particularly vocal cues, emotions reveal themselves to the outside observer.

While – judging from the remarkable ease and accuracy with which most individuals crack the emotional code embedded in the human voice – one might consider the decoding of emotional vocal information a simple matter; the neurocognitive mechanisms that mediate our abilities to derive emotional meaning from voice signals prove anything but simple: At the cerebral level, emotional voice perception draws upon the complex interplay of several cortical as well as subcortical brain structures. Years of research have led to identify superior temporal as well as dorsolateral frontal, orbitofrontal and limbic brain areas as constituents of a cerebral network implicated in the processing of emotional vocal information. Each structure of the network has been tied to distinct sub-processes or cognitive operations ranging from basic stages of acoustic analysis to higher-order evaluative processes: Whereas superior temporal brain regions have been suggested to subserve fundamental stages of acoustic analysis and perceptual binding, frontal structures (DLPFC/OFC) have been related to sub-processes governing the explicit evaluation of vocally expressed emotions and (social) decision making.

Although research of the past decades has greatly broadened our insight into the neurobiology of emotional voice perception, presented empirical evidence and models mark the beginning not the end of an endeavor to understand how the brain interprets emotional vocal cues. Rather than providing an exhaustive and coherent account, current knowledge forms a solid groundwork upon which to base future research needed to clarify, among others, several important issues outlined in this chapter that remain (as of yet) understudied.

References

[1] Russell JA, Bachorowski JA, Fernandez-Dols JM. Facial and vocal expressions of emotion. Annu Rev Psychol 2003;54:329–49.

[2] Johnstone T, Scherer KR. Vocal communication of emotion. In: Lewis M, Haviland J, editors. Handbook of emotion. New York: Guilford; 2000.

[3] Scherer KR, Johnstone T, Klasmeyer G. Vocal expression of emotion. In: Davidson RJ, Scherer KR, Goldsmith H, editors. Handbook of affective science. New York, Oxford: Oxford University Press; 2003.

[4] Pittam J, Scherer KR. Vocal expression and communication of emotion. In: Lewis M, Haviland JM, editors. Handbook of emotions. New York, London: The Guilford Press; 1993.

[5] Morton ES. On the occurrence and significance of motivation-structure rules in some bird and mammal sounds. American Naturalist 1977;111(981):855–69.

[6] Grandjean D, Banziger T, Scherer KR. Intonation as an interface between language and affect. Prog Brain Res 2006;156:235–47.

[7] Scherer U, Helfrich H, Scherer KR. Paralinguistic behaviour: internal push or external pull? In: Giles H, Robinson P, Smith P, editors. Language: social psychological perspectives. Oxford: Pergamon; 1980. p. 279–82.

[8] Scherer KR. Vocal affect expression: a review and a model for future research. Psychol Bull 1986;99(2):143–65.

[9] Bachorowski JA, Owren MJ. Sounds of emotion: production and perception of affect-related vocal acoustics. Ann N Y Acad Sci 2003;1000:244–65.

[10] Bachorowski JA. Vocal expression and perception of emotion. Curr Dir Psychol Sci 1999;8(53):53–7.

[11] Banse R, Scherer KR. Acoustic profiles in vocal emotion expression. J Pers Soc Psychol 1996;70(3):614–36.

[12] Juslin PN, Laukka P. Emotional expression in speech and music: evidence of cross-modal similarities. Ann N Y Acad Sci 2003;1000:279–82.

[13] Scherer KR, et al. Vocal cues in emotion encoding and decoding. Motivation and Emotion 1991;15(2):123–48.

[14] Szameitat DP, et al. Acoustic profiles of distinct emotional expressions in laughter. J Acoust Soc Am 2009;126(1):354–66.

[15] Sauter DA, et al. Perceptual cues in nonverbal vocal expressions of emotion. Q J Exp Psychol (Colchester) 2010;63(11):2251–72.

[16] Szameitat DP, et al. Acoustic correlates of emotional dimensions in laughter: arousal, dominance and valence. Cogn Emotion 2011;25(4):599–611.

[17] Belin P, Fecteau S, Bedard C. Thinking the voice: neural correlates of voice perception. Trends Cogn Sci 2004;8(3):129–35.

[18] Latinus M, Belin P. Human voice perception. Curr Biol 2011;21(4):R143–5.

[19] Grossmann T, et al. The developmental origins of voice processing in the human brain. Neuron 2010;65(6):852–8.

[20] Belin P, et al. Voice-selective areas in human auditory cortex. Nature 2000;403(6767):309–12.

[21] Belin P, Zatorre RJ. Adaptation to speaker's voice in right anterior temporal lobe. Neuroreport 2003;14(16):2105–9.

[22] Fecteau S, et al. Is voice processing species-specific in human auditory cortex? An fMRI study. NeuroImage 2004;23(3):840–8.

[23] Belin P, Zatorre RJ, Ahad P. Human temporal-lobe response to vocal sounds. Brain Res Cogn Brain Res 2002;13(1):17–26.

[24] Belin P, Grosbras MH. Before speech: cerebral voice processing in infants. Neuron 2010;65(6):733–5.

[25] Wiethoff S, et al. Cerebral processing of emotional prosody – influence of acoustic parameters and arousal. NeuroImage 2008;39(2):885–93.

[26] Ethofer T, et al. Decoding of emotional information in voice-sensitive cortices. Curr Biol 2009;19(12):1028–33.

[27] Ethofer T, et al. Differential influences of emotion, task, and novelty on brain regions underlying the processing of speech melody. J Cogn Neurosci 2009;21(7):1255–68.

[28] Ethofer T, et al. Effects of prosodic emotional intensity on activation of associative auditory cortex. Neuroreport 2006;17(3):249–53.

[29] Grandjean D, et al. The voices of wrath: brain responses to angry prosody in meaningless speech. Nat Neurosci 2005;8(2):145–6.

[30] Ethofer T, et al. The voices of seduction: cross-gender effects in processing of erotic prosody. Soc Cogn Affect Neurosci 2007;2(4):334–7.

[31] Scherer KR. Comment: Interpersonal expectations, social influence, and emotion transfer. In: Blanck PD, editor. Interpersonal expectations: theory, research and applications. Cambridge University Press & Editions de la Maison des Sciences de l'Homme Paris; 1993.

[32] Planalp S. Communicating emotion in everyday life: cues, channels, and processes. In: Andersen PA, editor. Handbook of communication and emotion: research, theory, application, and contexts. San Diego: Academic Press; 1998.

[33] Scherer KR, Banse R, Wallbott HG. Emotion inferences from vocal expression correlate across languages and cultures. J Cross Cult Psychol 2001;32(1):76–92.

[34] Pell MD, et al. Recognizing emotions in a foreign language. J Nonverbal Behav 2009;33(2):107–20.

[35] Ross ED. The aprosodias. Functional-anatomic organization of the affective components of language in the right hemisphere. Arch Neurol 1981;38(9):561–9.

[36] Broca PR. Sur le siège de la faculté du language articulé, suivies d'une observation d'aphémie (perte de la parole). Bull Soc Anat Paris 1861;36:330–57.

[37] Wernicke C. Der aphasische Symptomencomplex. Eine psychologische Studie auf anatomischer Basis. Breslau: Cohn & Weigert; 1874.

[38] Ackermann H, et al. Das Hören von Gefühlen: Funktionell-neuroanatomische Grundlage der Verarbeitung affektiver Prosodie. Akt Neurol 2004;31:449–60.

[39] Wildgruber D, et al. A cerebral network model of speech prosody comprehension. Int J Speech Lang Pathol 2009;11(4):277–81.

[40] Wildgruber D, et al. Cerebral processing of linguistic and emotional prosody: fMRI studies. Prog Brain Res 2006;156:249–68.

[41] Schirmer A, Kotz SA. Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. Trends Cogn Sci 2006;10(1):24–30.

[42] Ethofer T, et al. Emotional voice areas: anatomic location, functional properties, and structural connections revealed by combined fMRI/DTI. Cereb Cortex 2011 [e-pub ahead of print].

[43] Ethofer T, et al. Cerebral pathways in processing of affective prosody: a dynamic causal modeling study. NeuroImage 2006;30(2):580–7.

[44] Mitchell RL, et al. The neural response to emotional prosody, as revealed by functional magnetic resonance imaging. Neuropsychologia 2003;41(10):1410–21.

[45] Quadflieg S, et al. Modulation of the neural network involved in the processing of anger prosody: the role of task-relevance and social phobia. Biol Psychol 2008;78(2):129–37.

[46] Wildgruber D, et al. Distinct frontal regions subserve evaluation of linguistic and emotional aspects of speech intonation. Cereb Cortex 2004;14(12):1384–9.

[47] Wildgruber D, et al. Identification of emotional intonation evaluated by fMRI. NeuroImage 2005;24(4):1233–41.

[48] Hornak J, et al. Changes in emotion after circumscribed surgical lesions of the orbitofrontal and cingulate cortices. Brain 2003;126(Pt 7):1691–712.

[49] Hornak J, Rolls ET, Wade D. Face and voice expression identification in patients with emotional and behavioural changes following ventral frontal lobe damage. Neuropsychologia 1996;34(4):247–61.

[50] Baddeley A. Recent developments in working memory. Curr Opin Neurobiol 1998;8(2):234–8.

[51] Chein JM, Ravizza SM, Fiez JA. Using neuroimaging to evaluate models of working memory and their implications for language processing. J Neuroling 2003;16:315–39.

[52] D'Esposito M, et al. Functional MRI studies of spatial and nonspatial working memory. Brain Res Cogn Brain Res 1998;7(1):1–13.

[53] Smith EE, Jonides J. Storage and executive processes in the frontal lobes. Science 1999;283(5408):1657–61.

[54] Mitchell RL. fMRI delineation of working memory for emotional prosody in the brain: commonalities with the lexico-semantic emotion network. NeuroImage 2007;36(3):1015–25.

[55] Blair RJ, et al. Dissociable neural responses to facial expressions of sadness and anger. Brain 1999;122(Pt 5):883–93.

[56] Small DM, et al. Changes in brain activity related to eating chocolate: from pleasure to aversion. Brain 2001;124(Pt 9):1720–33.

[57] Kringelbach ML, Rolls ET. The functional neuroanatomy of the human orbitofrontal cortex: evidence from neuroimaging and neuropsychology. Prog Neurobiol 2004;72(5):341–72.

[58] Kanwisher N, McDermott J, Chun MM. The fusiform face area: a module in human extrastriate cortex specialized for face perception. J Neurosci 1997;17(11):4302–11.

[59] O'Craven KM, Kanwisher N. Mental imagery of faces and places activates corresponding stimulus-specific brain regions. J Cogn Neurosci 2000;12(6):1013–23.

[60] Wiethoff S, et al. Response and habituation of the amygdala during processing of emotional prosody. Neuroreport 2009;20(15):1356–60.

[61] Morris JS, Scott SK, Dolan RJ. Saying it with feeling: neural responses to emotional vocalizations. Neuropsychologia 1999;37(10):1155–63.

[62] Adolphs R, Tranel D. Intact recognition of emotional prosody following amygdala damage. Neuropsychologia 1999;37(11):1285–92.

[63] Bach DR, et al. The effect of appraisal level on processing of emotional prosody in meaningless speech. NeuroImage 2008;42(2):919–27.

[64] Ethofer T, et al. Impact of voice on emotional judgment of faces: an event-related fMRI study. Hum Brain Mapp 2006;27(9):707–14.

[65] Mitchell DG, et al. The impact of processing load on emotion. NeuroImage 2007;34(3):1299–309.

[66] Blair KS, et al. Modulation of emotion by cognition and cognition by emotion. NeuroImage 2007;35(1):430–40.

[67] Critchley H, et al. Explicit and implicit neural mechanisms for processing of social information from facial expressions: a functional magnetic resonance imaging study. Hum Brain Mapp 2000;9(2):93–105.

[68] Hariri AR, et al. Neocortical modulation of the amygdala response to fearful stimuli. Biol Psychiatry 2003;53(6):494–501.

[69] Tamietto M, de Gelder B. Neural bases of the non-conscious perception of emotional signals. Nat Rev Neurosci 2010;11(10):697–709.

[70] Sander D, et al. Emotion and attention interactions in social cognition: brain regions involved in processing anger prosody. NeuroImage 2005;28(4):848–58.

[71] Szameitat DP, et al. It is not always tickling: distinct cerebral responses during perception of different laughter types. NeuroImage 2010;53(4):1264–71.

[72] Amodio DM, Frith CD. Meeting of minds: the medial frontal cortex and social cognition. Nat Rev Neurosci 2006;7(4):268–77.

[73] Frith C, Frith U. Theory of mind. Curr Biol 2005;15(17):R644–6.

[74] LeDoux J. The emotional brain: the mysterious underpinnings of emotional life. London: Phoenix; 1998.

[75] Scott S, Caird FI, Williams BO. Evidence for an apparent sensory speech disorder in Parkinson's disease. J Neurol Neurosurg Psychiatry 1984;47(8):840–3.

[76] Dara C, Monetta L, Pell MD. Vocal emotion processing in Parkinson's disease: reduced sensitivity to negative emotions. Brain Res 2008;1188:100–11.

[77] Pell MD, Leonard CL. Processing emotional tone from speech in Parkinson's disease: a role for the basal ganglia. Cogn Affect Behav Neurosci 2003;3(4):275–88.

[78] Pell MD. On the receptive prosodic loss in Parkinson's disease. Cortex 1996;32(4):693–704.

[79] Brück C, et al. Effects of subthalamic nucleus stimulation on emotional prosody comprehension in Parkinson's disease. PLoS One 2011;6(4):e19140.

[80] Starkstein SE, et al. Neuropsychological and neuroradiologic correlates of emotional prosody comprehension. Neurology 1994;44(3 Pt 1):515–22.

[81] Cancelliere AE, Kertesz A. Lesion localization in acquired deficits of emotional expression and comprehension. Brain Cogn 1990;13(2):133–47.

[82] Paulmann S, Ott DVM, Kotz SA. Emotional speech perception unfolding in time: the role of the basal ganglia. PLoS One 2011;6(3):e17694.

[83] Paulmann S, Pell M, Kotz SA. Emotional prosody recognition in BG-patients: disgust recognition revisited. Brain Lang 2005;95:143–4.

[84] Kotz SA, et al. On the lateralization of emotional prosody: an event-related functional MR investigation. Brain Lang 2003;86(3):366–76.

[85] Breitenstein C, et al. Impaired perception of vocal emotions in Parkinson's disease: influence of speech time processing and executive functioning. Brain Cogn 2001;45(2):277–314.

[86] Brown LL, Schneider JS, Lidsky TI. Sensory and cognitive functions of the basal ganglia. Curr Opin Neurobiol 1997;7(2):157–63.

[87] Blonder LX, Gur RE, Gur RC. The effects of right and left hemiparkinsonism on prosody. Brain Lang 1989;36(2):193–207.

[88] Woods DL. The component structure of the N1 wave of the human auditory evoked potential. Electroencephalogr Clin Neurophysiol Suppl 1995;44:102–9.

[89] Wambacq IJ, Shea-Miller KJ, Abubakr A. Non-voluntary and voluntary processing of emotional prosody: an event-related potentials study. Neuroreport 2004;15(3):555–9.

[90] Scott SK, Sauter DA, McGettigan C. Brain mechanisms for processing perceived emotional vocalizations in humans. In: Brudzynski SM, editor. Handbook of mammalian vocalization an integrative neuroscience approach. Oxford: Academic Press; 2009. p. 187–98.

[91] Provine RR. Laughter: a scientific investigation. New York: Viking Books; 2000.

[92] Panksepp J. The riddle of laughter: neural and psychoevolutionary underpinnings of joy. Curr Dir Psychol Sci 2000;9(6):183–6.

[93] Meyer M, et al. Comparative evidence from behavioral, electrophysiological and neuroimaging studies in human and monkey. Behav Brain Res 2007;182:245–60.

[94] Szameitat DP, et al. Formant characteristics of human laughter. J Voice 2010.

[95] Szameitat DP, et al. Differentiation of emotions in laughter at the behavioral level. Emotion 2009;9(3):397–405.

[96] Meyer M, et al. Distinct fMRI responses to laughter, speech, and sounds along the human peri-sylvian cortex. Cogn Brain Res 2005;24(2):291–306.

[97] Sander K, Scheich H. Auditory perception of laughing and crying activates human amygdala regardless of attentional state. Cogn Brain Res 2001;12:181–98.

[98] Sander K, Scheich H. Left auditory cortex and amygdala, but right insula dominance for human laughing and crying. J Cogn Neurosci 2005;17:1519–31.

[99] Kreifelts B, et al. Audiovisual integration of emotional signals in voice and face: an event-related fMRI study. NeuroImage 2007;37(4):1445–56.

[100] de Gelder B, Vroomen J. The perception of emotions by ear and by eye. Cogn Emotion 2000;14(3):289–311.

[101] Massaro DW, Egan PB. Perceiving affect from the voice and the face. Psychonomic Bull Rev 1996;3(2):215–21.

[102] Dolan RJ, Morris JS, de Gelder B. Crossmodal binding of fear in voice and face. Proc Natl Acad Sci USA 2001;98(17):10006–10.

[103] Ethofer T, Pourtois G, Wildgruber D. Investigating audiovisual integration of emotional signals in the human brain. Prog Brain Res 2006;156:345–61.

[104] Campanella S, Belin P. Integrating face and voice in person perception. Trends Cogn Sci 2007;11(12):535–43.

[105] Kreifelts B, Wildgruber D, Ethofer T. Audiovisual integration of emotional information from voice and face. In: Belin P, Campanella S, Ethofer T, editors. Integrating face and voice in person perception. Berlin: Springer [in press].

[106] Ghazanfar AA, Chandrasekaran C, Logothetis NK. Interactions between the superior temporal sulcus and auditory cortex mediate dynamic face/voice integration in rhesus monkeys. J Neurosci 2008;28(17):4457–69.

[107] Ghazanfar AA, et al. Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. J Neurosci 2005;25(20):5004–12.

[108] Kayser C, Petkov CI, Logothetis NK. Multisensory interactions in primate auditory cortex: fMRI and electrophysiology. Hear Res 2009;258(1–2):80–8.

[109] Damasio AR. Time-locked multiregional retroactivation: a systems-level proposal for the neural substrates of recall and recognition. Cognition 1989;33(1–2):25–62.

[110] Jones EG, Powell TP. An anatomical study of converging sensory pathways within the cerebral cortex of the monkey. Brain 1970;93(4):793–820.

[111] Seltzer B, Pandya DN. Afferent cortical connections and architectonics of the superior temporal sulcus and surrounding cortex in the rhesus monkey. Brain Res 1978;149(1):1–24.

[112] Chavis DA, Pandya DN. Further observations on corticofrontal connections in the rhesus monkey. Brain Res 1976;117(3):369–86.

[113] Mesulam MM, Mufson EJ. Insula of the old world monkey. III: Efferent cortical output and comments on function. J Comp Neurol 1982;212(1):38–52.

[114] Fries W. Cortical projections to the superior colliculus in the macaque monkey: a retrograde study using horseradish peroxidase. J Comp Neurol 1984;230(1):55–76.

[115] Pearson RC, et al. The organization of the connections between the cortex and the claustrum in the monkey. Brain Res 1982;234(2):435–41.

[116] Mufson EJ, Mesulam MM. Thalamic connections of the insula in the rhesus monkey and comments on the paralimbic connectivity of the medial pulvinar nucleus. J Comp Neurol 1984;227(1):109–20.

[117] McDonald AJ. Cortical pathways to the mammalian amygdala. Prog Neurobiol 1998;55(3):257–332.

[118] Kreifelts B, et al. Association of trait emotional intelligence and individual fMRI-activation patterns during the perception of social signals from voice and face. Hum Brain Mapp 2010;31(7):979–91.

[119] Kreifelts B, et al. Cerebral representation of non-verbal emotional perception: fMRI reveals audiovisual integration area between voice- and face-sensitive regions in the superior temporal sulcus. Neuropsychologia 2009;47(14):3059–66.

[120] Posamentier MT, Abdi H. Processing faces and facial expressions. Neuropsychol Rev 2003;13(3):113–43.

[121] Mitchell RLC. How does the brain mediate interpretation of incongruent auditory emotions? The neural responses to prosody in the presence of conflicting lexico-semantic cues. Eur J Neurosci 2006;24:3611–8.

[122] Wittfoth M, et al. On emotional conflict: interference resolution of happy and angry prosody reveals valence-specific effects. Cereb Cortex 2010;20(2):383–92.

[123] Beaucousin V, et al. FMRI study of emotional speech comprehension. Cereb Cortex 2007;17(2):339–52.

[124] Herbert C, et al. Amygdala activation during reading of emotional adjectives – an advantage for pleasant content. Soc Cogn Affect Neurosci 2009;4(1):35–49.

[125] Hamann S, Mao H. Positive and negative emotional verbal stimuli elicit activity in the left amygdala. Neuroreport 2002;13(1):15–9.

[126] George MS, et al. Understanding emotional prosody activates right hemisphere regions. Arch Neurol 1996;53(7):665–70.

[127] Sammler D, et al. Prosody meets syntax: the role of the corpus callosum. Brain 2010;133(9):2643–55.

[128] Hamann S, Canli T. Individual differences in emotion processing. Curr Opin Neurobiol 2004;14(2):233–8.

[129] Dalgleish T. The emotional brain. Nat Rev Neurosci 2004;5(7):583–9.

[130] Schirmer A, et al. Gender differences in the activation of inferior frontal cortex during emotional speech perception. NeuroImage 2004;21(3):1114–23.

[131] Schirmer A, Kotz SA, Friederici AD. Sex differentiates the role of emotional prosody during word processing. Brain Res Cogn Brain Res 2002;14(2):228–33.

[132] Brück C, et al. Impact of personality on the cerebral processing of emotional prosody. NeuroImage 2011;58(1):259–68.

[133] Borkenau P, Ostendorf F. NEO-FFI NEO-Fünf-Faktoren-Inventar nach Costa und McCrae. 2. Neu normierte und vollständig überarbeitet Auflage. Göttingen: Hogrefe; 2008.

[134] Hoekert M, et al. Impaired recognition and expression of emotional prosody in schizophrenia: review and meta-analysis. Schizophr Res 2007;96(1–3):135–45.

[135] Emerson CS, Harrison DW, Everhart DE. Investigation of receptive affective prosodic ability in school-aged boys with and without depression. Neuropsychiatry Neuropsychol Behav Neurol 1999;12(2):102–9.

[136] Kan Y, et al. Recognition of emotion from moving facial and prosodic stimuli in depressed patients. J Neurol Neurosurg Psychiatry 2004;75:1667–71.

[137] Murphy D, Cutting J. Prosodic comprehension and expression in schizophrenia. J Neurol Neurosurg Psychiatry 1990;53(9):727–30.

[138] Uekermann J, et al. Perception of affective prosody in major depression: a link to executive functions? J Int Neuropsychol Soc 2008;14(4):552–61.

[139] Chevallier C, et al. What's in a voice? Prosody as a test case for the theory of mind account of autism. Neuropsychologia 2011;49(3):507–17.

[140] Golan O, Baron-Cohen S, Hill JJ. The Cambridge mindreading (CAM) face–voice battery: testing complex emotion recognition in adults with and without Asperger syndrome. J Aut Dev Disord 2006;36(2):169–83.

[141] Golan O, et al. The 'reading the mind in the voice' test-revised: a study of complex emotion recognition in adults with and without autism spectrum conditions. J Aut Dev Disord 2007;37:1096–106.

[142] Van Lancker D, Cornelius C, Kreiman J. Recognition of emotional-prosodic meaning in speech by autistic, schizophrenic, and normal children. Dev Neuropsychol 1989;5(2):207–26.

[143] Bach DR, et al. Altered lateralisation of emotional prosody processing in schizophrenia. Schizophr Res 2009;110(1–3):180–7.

[144] Mitchell RL, et al. Neural response to emotional prosody in schizophrenia and in bipolar affective disorder. Br J Psychiatry 2004;184:223–30.

[145] Gervais H, et al. Abnormal cortical voice processing in autism. Nature Neurosci 2004;7(8):801–2.

[146] Leitman DI, et al. The neural substrates of impaired prosodic detection in schizophrenia and its sensorial antecedents. Am J Psychiatry 2007;164(3):474–82.

[147] Frommann N, Streit M, Wölwer W. Remediation of facial affect recognition impairments in patients with schizophrenia: a new training program. Psychiatry Res 2003;117(3):281–4.

[148] Wölwer W, et al. Remediation of impairments in facial affect recognition in schizophrenia: efficacy and specificity of a new training program. Schizophr Res 2005;80(2–3):295–303.

[149] Habel U, et al. Training of affect recognition in schizophrenia: neurobiological correlates. Soc Neurosci 2010;5(1):92–104.